Tn7 TRANSPOSES INTO REPLICATING DNA USING AN INTERACTION WITH THE PROCESSIVITY FACTOR, FACILITATING GENOME EVOLUTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Adam Robert Parks

August 2008

Tn7 TRANSPOSES INTO REPLICATING DNA USING AN INTERACTION WITH THE PROCESSIVITY FACTOR, FACILITATING GENOME EVOLUTION

Adam Robert Parks, Ph. D.

Cornell University 2008

Transposons are mobile genetic elements that can move between locations in DNA that lack homology. Transposons play an important role in the evolution of genomes in every domain of life through activities such as horizontal gene transfer, gene disruption, gene expression modulation, and recombination. The bacterial transposon Tn7 maintains two distinct lifestyles, one in horizontally transferred DNA and the other in bacterial chromosomes. Access to these two DNA pools is mediated by two separate target selection pathways. The proteins involved in these pathways have evolved to specifically recognize their cognate target-sites using entirely different mechanisms but the same core transposition machinery. In this work, I analyze over 50 Tn7-like transposons and discuss how the molecular mechanisms of these genetic elements contribute to the success of both transposon and host. I focus particularly on the TnsE-pathway of transposition that is credited with optimizing transposition into DNA that is transported between bacterial hosts, and likely explains the presence of this transposon in phylogenetically diverse bacteria occupying a broad range of ecological niches. I show that the TnsE protein physically and functionally interacts with the processivity factor of the DNA replication machinery. I propose that this interaction allows Tn7 to identify insertion sites and to orient in one direction

with active DNA replication by a process that is ubiquitous in bacteria. The TnsE interaction with an essential and conserved component of the replication machinery reveals a new mechanism by which Tn7, and possibly other elements, select target-sites associated with DNA replication. These results provide insight into various processes in eukaryotic and prokaryotic organisms involving processivity factors. I also analyze specific genetic pathways that affect the frequency of TnsE-mediated transposition, and correlate these genetic effects with protein-DNA complexes that may be recognized by TnsE and could be expected to be commonly found in these genetic backgrounds. The data presented here reveals ways in which Tn7 has directed the evolution of host genomes, and points to ways in which Tn7 might be used as a tool for understanding genetic phenomena.

BIOGRAPHICAL SKETCH

Adam Parks was born in rural northern California, where he spent the first twenty years of his life. His early education was conducted in a small school in a town called Middletown. This rustic setting instilled in him a deep and abiding appreciation for the elegance and beauty of the natural world. While at Santa Rosa Junior College, Adam was particularly encouraged by Dr. Robert Rubin, whose stories and analogies amused and awed Adam into reverence of Biological Science.

Adam worked various odd-jobs while slowly making his way through an undergraduate degree, and eventually transferred to Humboldt State University in Arcata, California, where he completed his Bachelors of Science degree. While working in the labs of Dr. Patricia Siering and Dr. Jeffery Schineller, Adam found scientific inquiry to be fascinating and rewarding. The communities that Adam encountered at Humboldt State University also brought social and environmental responsibility to Adam's attention.

Drawn by academic rigor and the high quality of scientific research, Adam moved to Ithaca to attend Cornell University. Once at Cornell, Adam settled in the laboratory of Dr. Joseph Peters, where he cultivated an interest in molecular genetics. During his time as a graduate student at Cornell, Adam and his wife Sophie had a child, Jasper. With the birth of his son, Adam's education was transformed from one of scientific detachment to the holistic and humanistic endeavor that encompasses raising a family in an academic environment.

This thesis is dedicated to my wife, Sophie, and my son, Jasper.

ACKNOWLEDGMENTS

I would like to acknowledge the people who have provided me with the immense amount of help and encouragement on my academic path. My parents, Richard and Dale Parks provided me with a safe and caring childhood and continue to encourage and support my efforts. My grandmother and grandfather, the late Robert and Carol Morse, fed my curiosity by exposing me to nature and literature. Dr. Robert Rubin of Santa Rosa Junior College was the first college professor who led me to believe that I could, and should, pursue an advanced degree in science. Dr. Patricia Siering and Dr. Jeffery Schineller at Humboldt State University gave me my first opportunities to conduct research in their labs, and gave me valuable advice regarding all things academic. I am deeply indebted to committee members Dr. Jeffery Roberts, Dr. Bik Tye and particularly my committee chair Dr. Joseph Peters. Their advice and insight has been invaluable. Joseph Peters has been a very good friend and understanding supervisor, who has given me the freedom to explore and served as a fine example of who and what a scientist ought to be. I am thankful to the insightful discussions and technical advice that my friends and colleagues in the lab have provided me with. Jose Carlos Huguet, Qiaojuan Shi, Zaoping Li, Jessica Finn, Ben Porter, and Ilan Safir have all been very helpful and encouraging throughout my time in the Peters lab. Finally, I am grateful and thankful for the two people that have supported me the most and kept it all so interesting: my wife Sophie, and son Jasper.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION[1]

Transposons are mobile genetic elements that can move between locations in DNA that lack homology. Transposons play an important role in the evolution of genomes in every domain of life through activities such as horizontal gene transfer, gene disruption, gene expression modulation, and recombination (Bushman, 2002; Kazazian, 2004; Osborn and Boltner, 2002). They are common in plasmids, integrative conjugal elements (ICEs), bacteriophages, and chromosomes, and can transfer between hosts by moving from stable chromosomal sites to mobile DNA molecules. Tn7 is a bacterial transposon that possesses two separate transposition targeting systems that take advantage of both the stability of the chromosome and the mobility of plasmids and bacteriophages for propagation, persistence, and dissemination amongst bacteria (Barth et al., 1976; Craig, 2002; Finn et al., 2007; Kubo and Craig, 1990; Peters and Craig, 2001a; Peters and Craig, 2001b; Waddell and Craig, 1988; Wolkow et al., 1996). In Chapter 2, I will focus on the advantages of interplay between the two transposition pathways found in Tn7-like elements. I propose that the ability to move specifically between mobile DNAs and a single neutral position found in all bacterial chromosomes offer unique benefits to both host and transposon. In Chapter 2, I will also describe the analysis of 50 Tn7-like elements, which are easily

---

[1] This chapter has been prepared as part of a review and submitted to the journal Plasmid

Parks, A. R., and Peters, J. E. (submitted). Tn7 elements: Engendering diversity from chromosomes to episomes. Plasmid.

distinguished from other transposons based on the homology of at least four unique transposition gene components, the end architecture composed of a distinctive arrays of transposase binding sites, and frequently the localization at a specific chromosomal location. Examining the make-up of these transposons suggests ways in which Tn7-like elements have diverged from one another, what components of their transposition machinery are conserved, and how they have affected the evolution of host genomes.

From a “gene-centric” perspective of evolution we can think about the specific advantages afforded a collection of genes that can move as a group in different ways (e.g. in the form of a plasmid, bacteriophage, or transposon). A collection of genes that define a plasmid gains the ability to easily “shop around” to find the best fit in a number of bacterial hosts adapted to different environments by existing as an autonomous entity. This is especially true if the plasmid is capable of actively mobilizing its transfer between hosts as a conjugal plasmid. However, a downside to this strategy is that plasmids also risk being lost from a host during cell division or digested by nucleases in a new host. Alternatively, a collection of genes found on the chromosome is not as easily lost as extra chromosomally-encoded genes, but they also lose the ability to easily sample new hosts for the best evolutionary fit in different environments. The programmed ability of a collection of genes to reside in the chromosome and mobilize to new hosts will afford that collection of genes the advantages of both strategies. Consistent with this idea, there are many examples in nature where mechanisms have evolved for maximizing the efficient transfer of a collection of genes between a neutral integration site in the chromosome and into a form capable of mobilizing between hosts. Bacteriophages and integrative conjugal elements (ICEs) are both able to

utilize integration sites in the chromosome, but also have the ability to move between bacteria. Bacteriophages will frequently have a single “attachment site” in the chromosome, but also the ability to form an infecting particle to transport their genes to new hosts. ICEs are collections of genes found integrated within the chromosome that can also recombine out of the chromosome as a circle and transfer by conjugation to new bacterial hosts (Osborn and Boltner, 2002). Bacteriophages and ICEs are widespread and are also credited for the formation of pathogenicity or fitness islands (also known as genomic islands) in bacterial hosts. Genomic islands are large regions of the chromosome (>30 Kb) that have originated from horizontal gene transfer and contain genes with fitness enhancing qualities (Dobrindt et al., 2004). There are two important downsides to the life strategy of bacteriophages and ICEs. For one, these elements often need to encode 30 or more genes to carry out their elaborate form of transport to new hosts. A second downside of the strategies of bacteriophages and ICEs is that each system of packaging or conjugation also comes with some limitation on host range. As we will see with the Tn7-like elements these problems are eliminated. Tn7 elements use four gene products to specifically target mobile plasmids and a subset of bacteriophages for transposition by recognizing a very generic, albeit incompletely understood, facet of DNA replication found with some forms of mobile DNA. Therefore, Tn7-like elements can access mobile DNAs with exceptional host-range and without the need to carry the genes for mobilization.

Tn7 is an intricate transposon that displays many innovations distinguishing it from other transposons. It has been extensively studied as a model genetic element and many of its activities have been described in great

detail. Since its discovery as a plasmid borne antibiotic resistance determinant (Barth et al., 1976; Hedges et al., 1972), Tn7 and related elements have been found in a wide variety of clinical and environmental settings (Biskri and Mazel, 2003; Oppon et al., 1998; Orman et al., 2002; Parks and Peters, 2007; Ramirez et al., 2005a; Ramirez et al., 2005b). The Tn7 element, originally called TnC, was isolated from a trimethoprim resistant *Escherichia coli* that had infected a calf just two years after the introduction of this synthetic antibiotic into use in veterinary settings (Hedges et al., 1972). The element was propagated on a self transmissible IncIα plasmid, known as R-factor R483, however it was quickly discovered that the transposon was capable of integrating into the host genome in a highly site-specific manner (Barth et al., 1976; Lichtenstein and Brenner, 1981; Lichtenstein and Brenner, 1982).

Tn7-like elements appear to carry extremely diverse arrays of genes. These elements reside in organisms in both natural environments and those modified by humans (i.e. medical, industrial, and agricultural settings) (Parks and Peters, 2007). As investigators steadily flood genome databases with new sequences, Tn7-like elements continue to appear in unique and unusual organisms and environments. The success of these elements may be in large part attributed to an extremely fine tuned regulation of transposition (Craig, 2002; Peters and Craig, 2001b). Tn7 possesses two target-site selection pathways that evaluate potential insertion sites before activating a core transposition machinery which carries out the biochemical steps involved in moving the element (Kubo and Craig, 1990; Waddell and Craig, 1988). One pathway has evolved to maximize the efficiency of vertical transmission of the element by directing transposition into the chromosome (Lichtenstein and

Brenner, 1981; Lichtenstein and Brenner, 1982), and the other pathway has been optimized for horizontal gene transfer by preferentially directing transposition into mobile, or conjugal, plasmids and filamentous bacteriophages (Finn et al., 2007; Wolkow et al., 1996). The regulation of Tn7 differs from many other elements in that transposition does not occur unless specific targets have first been identified (Bainton et al., 1993; Stellwagen and Craig, 1997b; Stellwagen and Craig, 1998).

1.1 General overview of Tn7 transposition mechanism

Of the five transposition-associated gene products produced by Tn7, three are involved in regulation of transposition, while the remaining two carryout the actual chemical steps involved in mobilizing the element (Bainton et al., 1993; Hauer and Shapiro, 1984; Waddell and Craig, 1988)(Figure 1.1). The “core transposition machinery” of Tn7 consists of the transposase proteins TnsA and TnsB along with an AAA regulator protein, TnsC (Bainton et al., 1993). This core machinery is directed by one of two target selecting proteins, TnsD or TnsE (Kubo and Craig, 1990; Waddell and Craig, 1988). Tn7 transposes via a cut-and-paste mechanism where the element is completely excised from a donor DNA site and reinserted into a new target DNA. However, transposition is only activated once a target has been identified, allowing a nucleoprotein complex with donor DNA, target DNA, and transposition proteins to be formed (Bainton et al., 1991; Bainton et al., 1993; Skelding et al., 2002).

1.2. The core transposition Machinery

TnsB is a DDE-type transposase that is a member of the retroviral

integrase superfamily (Andrake and Skalka, 1996; Sarnovsky et al., 1996). This protein carries out the concerted breakage and rejoining reactions, joining the 3'-OH of the donor ends to 5'-PO_4 groups at the insertion site of the target molecule (Sarnovsky et al., 1996). The TnsA protein structurally resembles a restriction endonuclease (Ronning et al., 2004) and carries out the nicking reaction on the opposite strand of the donor molecule, completely freeing the element and leaving behind a DNA double-strand break (DSB) (Bainton et al., 1991; May and Craig, 1996; Sarnovsky et al., 1996). The sites where the top and bottom

strands are joined to the target molecule are offset by 5 bp, which creates 5 bp gaps in the target DNA at either end of the newly inserted transposon. Repair of these gaps by host machinery after transposition results in a 5 bp duplication at the site of insertion that is characteristic of Tn7 transposition (Bainton et al., 1991).

Normally, TnsA and TnsB will not carryout transposition alone. TnsC modulates the activity of the TnsAB transposase, and only activates transposition once complexed with target DNA and one of the target selection proteins, TnsD or TnsE (Bainton et al., 1993; Kubo and Craig, 1990; Waddell and Craig, 1988). TnsC is a double-stranded DNA binding protein that is believed to modulate activity through an ATPase domain (Stellwagen and Craig, 1997b; Stellwagen and Craig, 1998). This protein is considered the central regulator of transposition due to its involvement in a process called target-site immunity (see below) and its requirement in both pathways of target-site selection (Waddell and Craig, 1988).

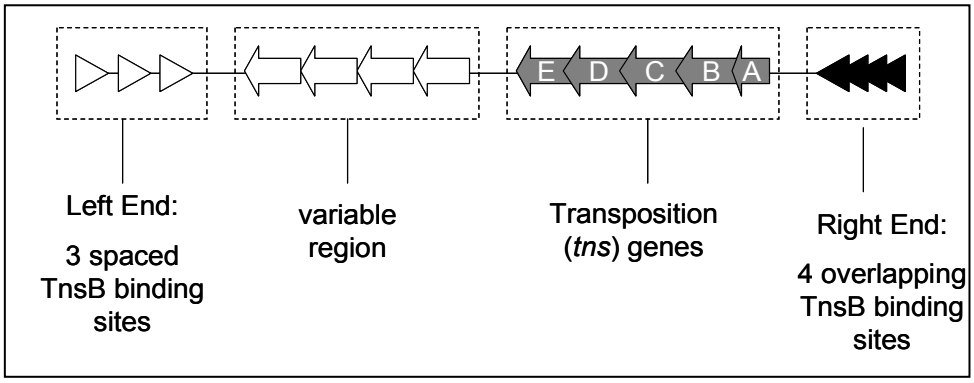


Figure 1.1. Generalized Tn7-like element diagram
Tn7-like elements display a high degree of diversity while maintaining conservation of components that are essential for propagation and dissemination of the element. This schematic representation of Tn7 elements shows the various components found in these transposons. The TnsB binding sites that delineate the ends of the transposon are shown as triangles. All DNA between these binding sites is mobilized as a single unit. Genes involved in transposition of the Tn7 element are shown as filled arrows. The variable region in the left end of the element is represented with open arrows.

The two target selecting proteins TnsD and TnsE act by identifying potential target molecules, activating the core machinery through the formation of the nucleoprotein complex, and directing transposition into those DNA molecules (Craig, 2002; Finn et al., 2007; Peters and Craig, 2001b). Specific protein-protein interactions between TnsC and target-selection proteins may not be necessary, as some TnsC mutants are responsive to certain alternative DNA structures (Rao et al., 2000).

The Tn7 right and left ends are distinct from one another in that they are composed of differing configurations of TnsB binding sites with varying affinities for the TnsB protein (Arciszewska et al., 1989; McKown et al., 1987) (Figure 1.1). The right end contains four overlapping TnsB binding sites, while the left end contains three widely spaced TnsB binding sites (McKown et al., 1987; Tang et al., 1991). This elaborate configuration of binding sites imparts an asymmetry upon the transposon and somehow allows the unique behavior of Tn7 to control not only the position and timing of transposition, but also the left-to-right orientation of transposition insertion events (Bainton et al., 1991; Finn et al., 2007; Gringauz et al., 1988; Lichtenstein and Brenner, 1982; McKown et al., 1988; Peters and Craig, 2001a). In the laboratory it has been shown that transposition can occur using two right ends, but not with two left ends (Arciszewska et al., 1989).

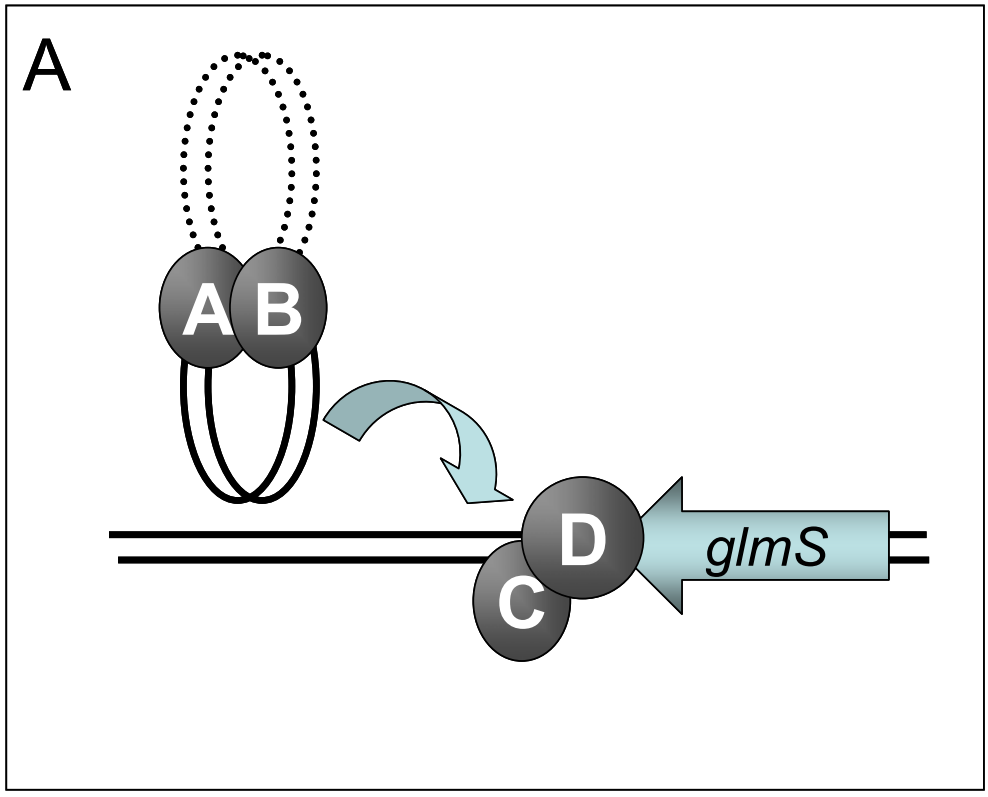
1.3. The TnsD-mediated pathway of transposition

Tn7 utilizes two distinct pathways for targeting transposition, one catalyzed by TnsD and the other catalyzed by TnsE (Kubo and Craig, 1990; Waddell and Craig, 1988). The TnsD protein recognizes and binds to a highly conserved DNA sequence in the 3' end of the *glmS* gene, directing insertions

into a single site within the *glmS* transcriptional terminator (Figure 1.2). The TnsD binding site within the *glmS* open reading frame and the actual site of Tn7 insertion within the transcriptional terminator are collectively referred to as the Tn7 attachment site, or *attTn7* (Figure 1.3.A.)(Gary et al., 1996; Lichtenstein and Brenner, 1982; McKown et al., 1988). Insertion into *attTn7* occurs at a very high frequency (~10^{-1}-10^{-2} chromosomal insertions per single copy donor plasmid)(DeBoy, 1997), causes no detectable deleterious effect, and the insertion event is stable even without selection (McKenzie and Craig, 2006).

Binding of TnsD to DNA causes a distortion in the duplex DNA structure that recruits TnsC to *attTn7*, activating transposition (Kuduvalli et al., 2001). Transposition into *attTn7*is very specific, almost always occurring at a single set of base pair junctions within a given organism with the right end of the element proximal to the end of the *glmS* open reading frame (DeBoy, 1997; Gringauz et al., 1988; Lichtenstein and Brenner, 1981; Lichtenstein and Brenner, 1982). Some organisms contain more than one *glmS* homolog and therefore more than one *attTn7* (Choi et al., 2006; Kuduvalli et al., 2005). Analysis of Tn7-like elements indicates that the distance between the site of insertion and the end of the *glmS* open reading frame varies slightly between organisms (Craig, 1989; Parks and Peters, 2007). The TnsD protein also interacts with two host proteins, ribosomal protein L29 and acyl-carrier protein (ACP), which appear to stabilize the nucleoprotein complex and stimulate TnsABC+D transposition (Sharpe and Craig, 1998). Interaction with these proteins may provide the Tn7 element with cues regarding host cell metabolic state or growth phase, further refining regulation of transposition (Sharpe and Craig, 1998).

Figure 1.2. Diagram describing TnsD-mediated transposition.
A. TnsD binds to the 3' end of the *glmS* gene, recruiting TnsC. The TnsC-TnsD-attTn7 complex activates transposition into the transcriptional terminator of the *glmS* gene.
B. TnsD binds to the major groove of DNA in the *glmS* gene between 30 and 55 (+33 and +55) bases away from the actual site of transposon insertion. TnsD binding causes a distortion of the DNA in the +27 region, allowing recognition and binding by TnsC. TnsC binds to the minor groove of the DNA, extending the footprint of the TnsD-TnsC complex up to the region of transposon insertion (0), serving as a platform for the transposase to access the target DNA. The 0 position is centered on the center base pair of the five base pairs that are duplicated as a result of transposon insertion. The ends of the transposon are represented by triangles, with the "L" representing the left end and "R" representing the right end. This figure is modified from Kuduvalli, et. al. (Kuduvalli et al., 2001).


A
A
B
D
C
glmS

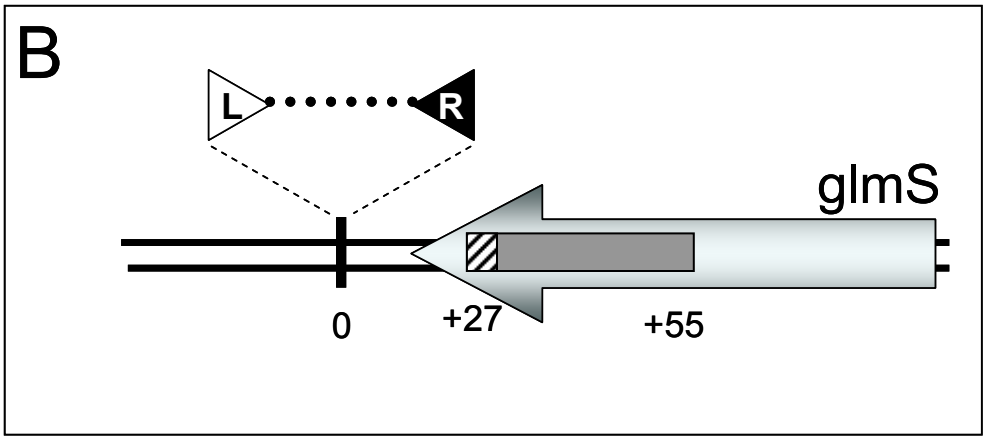
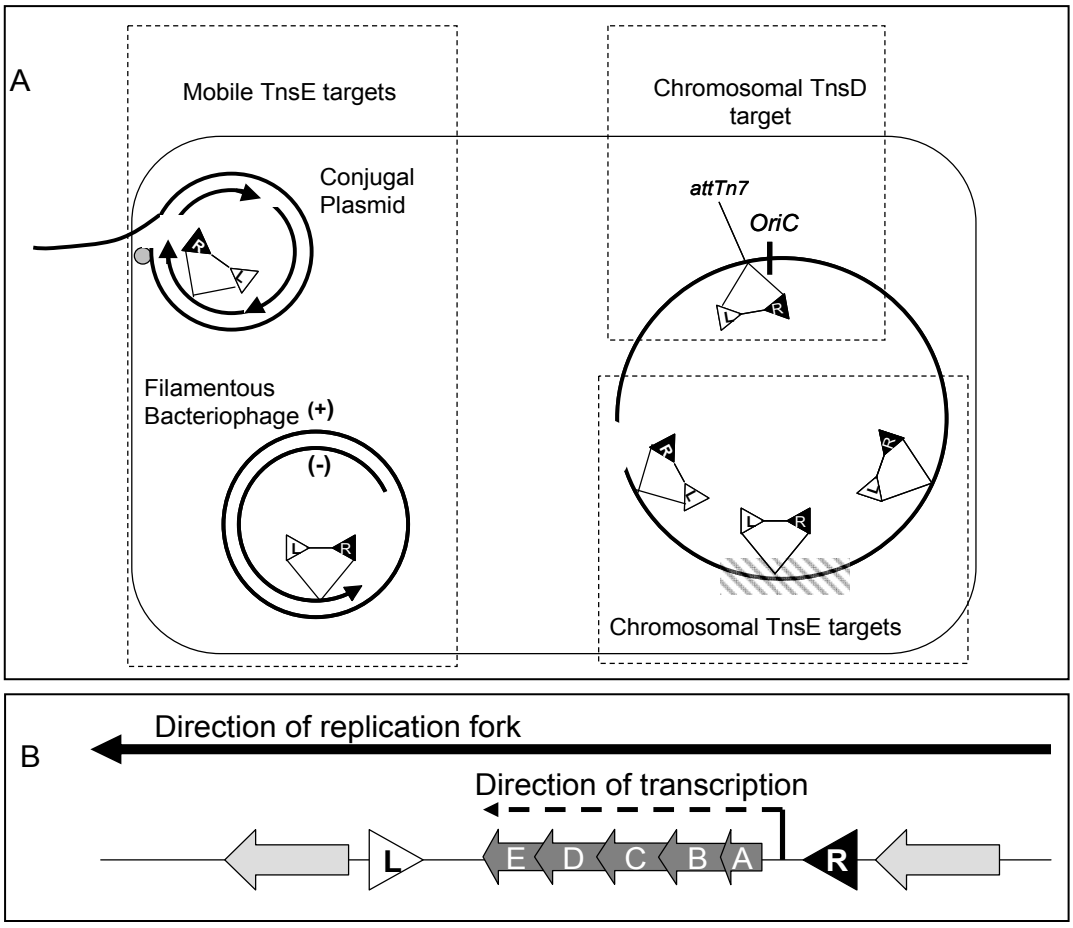

B
L
R
glmS
0
+27
+55

Figure 1.3. Target-site selection pathways and orientation of transposon insertion.
A. Schematic representation of TnsD and TnsE targets. TnsD insertions occur within a single site in the chromosome with a single left to right orientation ("L" represents left end, and "R" represents Right end). TnsE preferentially targets mobile DNA. Small circle represents the relaxase that covalently binds the 5' end of mobile plasmids, arrowheads represent 3' ends, and arrow tails represent 5' ends. The (+) and (-) distinguish the opposite strands of filamentous bacteriophage. Some TnsE targets can be found within the chromosome. Chromosomal targets include sites proximal to double-stranded DNA breaks (opening in chromosome), and the region of the chromosome where DNA replication terminates (crosshatched box). Orientation bias of TnsE-mediated insertions are reversed in each replichore (right and left side of the chromosome) and depends on the direction of the replication fork.
B. Both TnsD and TnsE mediate insertions that align the direction of transposon gene transcription with the direction of replication passage, minimizing head-to-head collisions with DNA replication machinery and transcription machinery (Sherratt, 2003).


A
Mobile TnsE targets
Chromosomal TnsD target
Conjugal Plasmid
attTn7
OriC
Filamentous Bacteriophage
(+)
(-)
Chromosomal TnsE targets
B
Direction of replication fork
Direction of transcription
L
E
D
C
B
A
R

The *glmS* gene is essential to the host and the primary amino acid sequence is well conserved in bacteria, making recognition of the coding sequence of this gene an elegant adaptation for Tn7. In addition, the chromosomal location of the *glmS* gene could also be considered advantageous for Tn7. Even in diverse species of bacteria the *glmS* gene is nearly always located close to the chromosomal origin of replication, and is therefore replicated early in the cell cycle. Since Tn7 is a cut-and-paste transposon, it leaves behind a DNA double-strand break (DSB) following excision (Bainton et al., 1991). A copy of the element is usually regenerated once the DSB is repaired by homologous recombination with the sister chromosome (Hagemann and Craig, 1993). The location of *attTn7* increases the probability that the DSB created by excision of the element is able to be repaired, and the transposon regenerated, even in slowly dividing cells. This observation is supported by work *in vivo* which showed that an interrupted *lac* operon near *attTn7* is restored to a functional Lac^+ status by homologous recombination with an adjacent defective *lac* operon following the excision of Tn7 (Hagemann and Craig, 1993).

Tn7-like elements can possess two distinct *tnsD* genes. No other *tns* genes appear to have been duplicated in a similar way, so it is likely that this duplication constitutes an adaptation that in some way benefits the transposon. For example, the auxiliary *tnsD* gene may more efficiently find attachment sites in phylogenetically distant organisms. Alternatively, unlike in Tn7, one of the TnsD proteins might actually allow non-specific target-site recognition. This would be consistent with the observation that elements encoding two distinct TnsD proteins are common in Tn7-like elements that are not found within the *attTn7* site (see Chapter 2). The incidence of elements

containing two distinct *tnsD* genes appears in ~7/50 elements (Chapter 2, Table 2.1). While it is possible that the additional TnsD proteins may identify attachment sites that have not yet been described, I find no evidence for a specific alternate attachment site in the available DNA sequences. There is no evidence of homology between TnsD and the other target-site selecting protein, TnsE, and therefore it appears unlikely that *tnsE* arose from tandem duplication of *tnsD*.

1.4. The TnsE-mediated pathway of transposition

While the TnsD pathway is exceptional at directing transposition into chromosomes and promoting vertical propagation of Tn7 to a host's progeny, the TnsE pathway enables horizontal transfer of the element (Finn et al., 2007; Wolkow et al., 1996). TnsE recognizes an entirely different class of target that is not at the level of nucleotide sequence (Finn et al., 2007; Peters and Craig, 2000; Peters and Craig, 2001a; Peters and Craig, 2001b; Shi, 2008a; Wolkow et al., 1996). Instead TnsE-mediated transposition occurs by identifying an aspect of lagging-strand DNA synthesis (Peters and Craig, 2001a; Peters and Craig, 2001b). This pathway of transposition preferentially recognizes mobile plasmids as they enter the Tn7 containing host cell (Wolkow et al., 1996) (Figure 1.2.A.). By targeting a process that occurs during conjugal replication within a new host, Tn7 gains access to DNA that is actively being transported amongst bacteria, therefore increasing the likelihood that the transposon will be transported to a new host. Tn7 is able to achieve this without the evolutionary trade off that other transposons face by selecting targets at random, risking the disruption of essential host functions. In Chapter 4, I address specific advantages that targeted transposition may present

compared to strictly random transposition.

It is the process of conjugal replication that activates plasmid DNAs as targets for TnsE-mediated transposition. Vegetative replication does not stimulate or preferentially attract transposition into plasmids via the TnsE pathway. However, mobilization of *oriT* containing plasmids by Tra functions supplied *in trans* is sufficient to activate transposition into these targets (Wolkow et al., 1996). The nature of conjugal DNA replication suggests that TnsE preferentially recognizes lagging-strand DNA replication (Peters and Craig, 2001a; Peters and Craig, 2001b). As conjugal plasmids enter a new host cell they are replicated by the host DNA replication machinery using a discontinuous process akin to lagging-strand DNA synthesis (Lawley et al., 2004). The major difference between vegetative replication of a plasmid and transfer associated replication is that leading- and lagging-strand DNA synthesis occur in two physically isolated locations. Leading-strand synthesis occurs in the donor cell as the transferred strand is displaced and transported into the recipient cell. Lagging-strand synthesis is carried out by the DNA replication machinery of the new recipient host cell (Lawley et al., 2004).

Stimulation of TnsE-mediated transposition in recipient cells and orientation of Tn7 ends following insertion within plasmids gave the first hint that discontinuous conjugal replication was the active target for TnsE (Wolkow et al., 1996). Analysis of transposition within the chromosome strongly supported the view that the target of the TnsE pathway is lagging-strand DNA synthesis (Peters and Craig, 2001a; Peters and Craig, 2001b). By using a process associated with certain types of DNA replication, rather than a specific DNA sequence, the TnsE pathway allows Tn7 to direct transposition into a

wide variety of DNAs that are transported amongst cells, effectively broadening the host range of the transposon.

Half of all the TnsE-mediated insertions isolated from actively conjugating plasmids were shown to insert in a single orientation near the *oriT* locus, within 1-2 Kb of the 5' end that initially entered the cell (Peters and Craig, 2001a; Wolkow et al., 1996). Insertion close to *oriT* is observed in naturally isolated Tn7 insertions such as that found in the R721 (4 Kb between *oriT* the right end of Tn7, AP002527). By inserting proximal to the origin of transfer, Tn7 may benefit by placing itself on a segment of DNA that enters new host cells early, allowing early expression of transposon genes. Transposition into the *attTn7* locus on the chromosome could occur as soon as possible to escape host defenses, such as DNA restriction systems, that are activated by the incoming plasmid.

There are examples of Tn7-like elements that appear in plasmids that lack a *tnsE* gene. A notable example can be found within *Escherichia coli* APEC O1, pAPEC-O1-R (DQ517526), and *Serratia marcescens*, R478 (BX664015)(Gilmour et al., 2004; Johnson et al., 2006). All components that are necessary for transposition of the element appear to be intact, but there is no detectable *tnsE* gene. Since TnsE actively targets transposition to conjugal plasmids, these insertions may have resulted from TnsE proteins produced *in trans* from a gene encoded by a similar, more complete, Tn7-like element. Another explanation is that these elements ended up in plasmids through random transposition. Both elements contain a mutation in TnsB that has previously been isolated in the laboratory allowing random transposition but still takes cues from target-site selection proteins (see Chapter 2). It will be interesting to know how different Tn7-like elements interact and if they are

able to mediate the excision and insertion of related elements.

TnsE is also capable of directing transposition into replicating filamentous bacteriophages (Finn et al., 2007). Analysis of the TnsE-mediated insertion events in bacteriophage M13 revealed that, similar to those found in mobile plasmids, they occurred almost exclusively in a single orientation. The orientation of Tn7 ends after insertion into M13 is consistent with recognition of minus-strand DNA replication. Minus-strand synthesis produces a DNA strand complementary to the incoming single-stranded bacteriophage genome to make the double-stranded (replicative) form of the bacteriophage (Finn et al., 2007). The efficiency and rapidity of TnsE-targeted transposition is illustrated by insertions that occur into bacteriophage M13. For transposition to occur, the target DNA molecule must be at least partially double-stranded, a condition that exists for a limited time in the lifecycle of a filamentous bacteriophage. No insertions into the M13 genome were detected using a mutant TnsABC machinery that is capable of random transposition, suggesting that a specific targeting pathway is required to achieve insertion into the short-lived duplex filamentous bacteriophage genome that represents a minority of total DNA within a cell (Finn et al., 2007).

TnsE-mediated insertions occasionally occur in the chromosome, albeit at a much lower frequency than observed in conjugal plasmids. These insertions are enriched in the region of the chromosome where replication terminates (Peters and Craig, 2000) (Figure 1.2.A). Naturally occurring Tn7-like transposons can be found in chromosomal locations that are consistent with TnsABC+E transposition (Parks and Peters, 2007). TnsE-mediated insertions found in the chromosome occur with a conspicuous left-to-right end orientation bias that is dependent on the direction of the replication fork that

has duplicated a given section of the chromosome (Peters and Craig, 2000; Peters and Craig, 2001a). All bacterial chromosomes are replicated bi-directionally by two replication forks that emanate from a single origin. In the case of circular bacterial genomes these replication forks meet at the terminus region, dividing the chromosome into two replichores (Sherratt, 2003). TnsE-mediated insertions occur with an opposite orientation of Tn7 ends in each replichore in a configuration that is consistent with the recognition of chromosomal lagging-strand DNA synthesis (Peters and Craig, 2001a).

The transposon left-to-right end orientation alignment with replication may constitute a biological adaptation. The orientation bias observed both with TnsE- and TnsD-mediated insertion positions the *tns* genes of Tn7 in such a way that the direction of their transcription does not oppose the direction of replication (figure 1.2.B). The orientation bias may therefore limit damaging head-on collisions between DNA replication machinery and transcription machinery. This feature of transposition may be more important in organisms such as *Bacillus*, where coordination of the directionality of transcription and replication appeared to be more stringently enforced than in organisms such as *E. coli* (Wang et al., 2007). Interestingly, all of the lambda-like phage that have been examined have also evolved to insert in one orientation where transcription of the large operons occur in the same direction as DNA replication of the host (Campbell, 2002). The orientation of open reading frames other than the *tns* genes within Tn7-like elements is less conserved, and sometimes is found opposing the direction of *tns* gene transcription. Observation of the orientation of open reading frames within these elements and measurement of their transcriptional activity may serve as an indicator of an organism's tolerance to transcription and replication

collisions. All open reading frames within the Tn7-like element found in the *Staphylococcus* plasmid pLEW6932 are oriented in the same direction (Parks and Peters, 2007), while those found in Tn7 from *E. coli* can be divergent from the *tns* genes (Craig, 2002).

TnsE-mediated transposition is stimulated by induced DNA double-strand breaks in the chromosome (Peters and Craig, 2000). In these cases insertion events occur proximal to the location of an induced DSB and tend to do so in hotspots that can be hundreds of kilobases distant from the initial break site (Shi, 2008b). TnsE-mediated transposition appears to recognize factors associated with the repair of the DSB, such as replication mediated repair (Shi, 2008). Transposition into chromosomes via the TnsE pathway is possible in *recA*$^-$ cells, indicating that homologous recombination is not essential for TnsE-mediated transposition (see Chapter 4)(Peters and Craig, 2001a). It is somewhat premature to speculate about what is recognized during DSB repair in the chromosome, given that the molecular complex recognized by TnsE during replication has not been defined; however, it is possible that the mechanism that recognizes DSB repair also facilitates Tn7 transposition into bacteriophage that utilize homologous recombination as an important step in their own replication. Homologous recombination is important for replication in many bacteriophages including lambda and T4 (Mosig, 1998; Smith, 1983; Weigel and Seitz, 2006). These elements have not been thoroughly examined for their ability to serve as targets for TnsE-mediated transposition, and may be interesting topics of future research.

The mechanism of TnsE target identification has yet to be completely solved. It is known that TnsE is a DNA binding protein that preferentially binds to DNA structures presenting a 3' recessed end, as might be expected in

incomplete Okazaki fragments (Peters and Craig, 2001a; Peters and Craig, 2001b; Wolkow et al., 1996). A specific interaction between TnsE and the processivity factor of DNA replication also appears to be instrumental in TnsE-mediated targe-site selection, and is the topic of Chapter 3. Where tested, randomly isolated mutations in TnsE that increase TnsABC+E transposition frequency display enhanced DNA binding ability (Peters and Craig, 2001a). Most of these mutations are located in the C-terminus of the TnsE, but one increased-activity mutation (M37I) can be found in the N-terminus. In Chapter 3, I present evidence that M37I may represent a class of mutation that alters protein-protein interactions between TnsE and the processivity factor of DNA replication.

As I will explain in Chapter 3, the β processivity factor of DNA replication is an essential host factor that aids in the identification of TnsE targets, just as L29 and ACP are involved TnsD-mediated insertion (Sharpe and Craig, 1998). The L29 and ACP interaction with TnsD presumably provides Tn7 with information regarding the growth phase or metabolic status of a new host cell. TnsE may interact with β in such a way that provides cues based on the replication status of a given DNA molecule. Interaction with β may explain why conjugal replication stimulates TnsE-mediated transposition to a greater degree than chromosomal DNA replication. The analysis of TnsE-mediated transposition events continues to be a useful tool for understanding the replication and propagation of mobile DNA and will likely provide additional information about the modes and progression of DNA replication of diverse biological entities.

1.5. Mutant forms of the transposition machinery are capable of non-targeted transposition and can circumvent target-site immunity

Sites that contain a copy of the Tn7 transposon exhibit a property known as target-site immunity that allows the element to sense the presence of a preexisting copy of the transposon and prevent reinsertion (Arciszewska et al., 1989). This principle is mediated by the ends of the transposon and by the TnsB and TnsC proteins. TnsB binds to the ends of the element and causes a redistribution of the TnsC protein away from these end sequences, presumably by stimulating the ATPase activity of TnsC (Arciszewska et al., 1989; DeBoy and Craig, 1996; Skelding et al., 2003; Stellwagen and Craig, 1997a). The ability of target-site immunity to discourage transposition can even be detected between sites that are >190 Kb apart (DeBoy and Craig, 1996). The efficiency of this process is significant, but not complete. *In vivo*, target-site immunity prevents ~90% of potential insertions from occurring in occupied target-sites (DeBoy and Craig, 1996). Mutations in TnsB and TnsC have been isolated in the laboratory that reduce the efficiency of target-site immunity up to 20-fold (Skelding et al., 2003; Stellwagen and Craig, 1997a).

Mutations in *tnsC* that allow the TnsABC machinery to work in the absence of the TnsD and TnsE proteins are categorized in two classes based on the ability to maintain target-site selection and target-site immunity (Stellwagen and Craig, 1997b). TnsC proteins with class I mutations catalyze transposition with TnsA and TnsB, and are still able to function in target-site immunity and are responsive to signals from the target selection proteins (i.e. they will not direct transposition into targets that already contain a Tn7 and are directed by TnsE or TnsD when present). Class II mutations allow transposition with TnsA and TnsB, but do not facilitate target-site immunity and

are unresponsive to signals from the target selection proteins. TnsC^{A225V}, a class I mutant, hydrolyzes ATP more slowly than wild-type, and may therefore remain in an “on” state and mediate Tn7 insertions into non-traditional, random, targets (Stellwagen and Craig, 1997b). Consistent with notion that the prolonged ATP bound state of TnsC is lock in the on conformation, TnsC containing the non-hydrolyzable ATP analog AMP-PNP displays increased DNA binding ability and exhibits decreased target-site immunity (Bainton et al., 1993; Stellwagen and Craig, 1997a). Insertions that occur via these mutant pathways lack any recognizable sequence specificity (Seringhaus et al., 2006; Stellwagen and Craig, 1997b). Altered or defective ATPase activity of TnsC could explain the existence of some Tn7-like transposons in chromosomal sites other than the *attTn7* site. Mutations in the *tnsA* and *tnsB* genes have also been shown to allow TnsABC, and in some cases TnsAB, transposition without the use of target-site selection proteins (Lu and Craig, 2000). However, the molecular details of how these mutants allow untargeted transposition are not entirely clear.

Mutations found in TnsB that abrogate immunity also impair TnsB-TnsC interaction but still allow transposition have been isolated in the laboratory (Skelding et al., 2003). TnsB mutations that allow the loss of target-site immunity may encourage the accumulation of multiple Tn7-like elements in *attTn7*, leading to the formation of genomic islands in the *attTn7* locus (see Chapter 2)(Parks and Peters, 2007). Since it is possible to maintain the ability to transpose and lose target-site immunity (Skelding et al., 2003), there may be a dynamic relationship between Tn7-like elements, preventing or allowing tandem insertion to occur depending on selective pressures.

1.6 Dissertation outline

In the following chapters, I will discuss the ways in which Tn7 interacts with the host genome and genome maintenance machinery, paying special attention to the TnsE pathway of target-site selection.

Chapter 2 will deal with the diversity and evolution of Tn7 and related elements. I describe the use of bioinformatic tools to examine the differences and similarities between Tn7 and Tn7-like elements, and draw comparisons between naturally occurring and laboratory generated mutations within these elements. I argue that flux between mobile and stationary DNA pools resulting from the unique interplay between the two target-site selection pathways of Tn7 may explain the diversity observed in Tn7-like transposons and contributes to the success of host organisms. Analysis of these Tn7-like elements suggests areas of future research.

In Chapter 3 I will focus on the physical and functional interaction between TnsE and the host DNA replication processivity factor, the β clamp. Through the use of both in vivo and in vitro assays, I show that the interaction between TnsE and β is essential for transposition via the TnsE pathway. I also describe the use of an in vivo SOS induction assay which can be used to examine interaction with the processivity factor in the presence of TnsE alone, and may be used to address questions regarding protein traffic on the processivity factor. I will propose a model for the identification of potential targets via the TnsE pathway in which the processivity factor plays a key role throughout the process of transposition.

In Chapter 4, I will explore the ways in which transposition via the TnsE pathway can inform our understanding of DNA repair processes. I will examine mutations in the host machinery that affect the target selection process of TnsE and explain how interpretation of these results can help us understand the molecular mechanisms of DNA damage repair.

In Chapter 5, I will summarize my findings and propose future lines of research. I will also discuss the implications of this work on the current understanding of DNA recombination, replication, and repair.

REFERENCES

Andrake, M. D., and Skalka, A. M. (1996). Retroviral Integrase, Putting the Pieces Together 10.1074/jbc.271.33.19633. J Biol Chem *271*, 19633-19636.

Arciszewska, L. K., Drake, D., and Craig, N. L. (1989). Transposon Tn7 *cis*-acting sequences in transposition and transposition immunity. J Mol Biol *207*, 35-52.

Bainton, R., Gamas, P., and Craig, N. L. (1991). Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell *65*, 805-816.

Bainton, R. J., Kubo, K. M., Feng, J.-N., and Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell *72*, 931-943.

Barth, P. T., Datta, N., Hedges, R. W., and Grinter, N. J. (1976). Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. J Bacteriol *125*, 800-810.

Biskri, L., and Mazel, D. (2003). Erythromycin esterase gene ere(A) is located in a functional gene cassette in an unusual class 2 integron. Antimicrob Agents Chemother *47*, 3326-3331.

Bushman, F. (2002). Lateral DNA transfer : mechanisms and consequences (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

Campbell, A. M. (2002). Preferential orientation of natural lambdoid prophages and bacterial chromosome organization. Theor Popul Biol *61*, 503-507.

Choi, K. H., DeShazer, D., and Schweizer, H. P. (2006). mini-Tn7 insertion in bacteria with multiple glmS-linked attTn7 sites: example Burkholderia mallei ATCC 23344. Nat Protoc *1*, 162-169.

Craig, N. (1989). Transposon Tn7. Mobile DNA, 211–225.

Craig, N. L. (2002). Tn7. In Mobile DNA II, M. Lambowitz-Alan, ed. (Washington, DC, ASM Press), pp. 423-456.

DeBoy, R., and Craig, N. L. (1996). Tn7 transposition as a probe of *cis* interactions between widely separated (190 Kilobases Apart) DNA sites in the *Escherichia coli* chromosome. J Bacteriol *178*, 6184-6191.

DeBoy, R. T. (1997) Transposon Tn7: target DNAs can modulate Tn7 insertion, Ph.D. Thesis., Johns Hopkins University., Baltimore, Maryland.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol *2*, 414-424.

Finn, J. A., Parks, A. R., and Peters, J. E. (2007). Transposon Tn7 directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. J Bacteriol *189*, 9122-9125.

Gary, P. A., Biery, M. C., Bainton, R. J., and Craig, N. L. (1996). Multiple DNA processing reactions underlie Tn7 transposition. J Mol Biol *257*, 301-316.

Gilmour, M. W., Thomson, N. R., Sanders, M., Parkhill, J., and Taylor, D. E. (2004). The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. Plasmid *52*, 182-202.

Gringauz, E., Orle, K., Orle, A., Waddell, C. S., and Craig, N. L. (1988). Recognition of *Escherichia coli attTn7* by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion. J Bacteriol *170*, 2832-2840.

Hagemann, A. T., and Craig, N. L. (1993). Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. Genetics *133*, 9-16.

Hauer, B., and Shapiro, J. A. (1984). Control of Tn7 transposition. Mol Gen Genet *194*, 149-158.

Hedges, R. W., Datta, N., and Fleming, M. P. (1972). R factors conferring resistance to trimethoprim but not sulphonamides. J Gen Microbiol *73*, 573-575.

Johnson, T. J., Wannemeuhler, Y. M., Scaccianoce, J. A., Johnson, S. J., and Nolan, L. K. (2006). Complete DNA sequence, comparative genomics, and prevalence of an IncHI2 plasmid occurring among extraintestinal pathogenic Escherichia coli isolates. Antimicrob Agents Chemother *50*, 3929-3933.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Kubo, K. M., and Craig, N. L. (1990). Bacterial transposon Tn7 utilizes two different classes of target sites. J Bacteriol *172*, 2774-2778.

Kuduvalli, P., Rao, J. E., and Craig, N. L. (2001). Target DNA structure plays a critical role in Tn7 transposition. EMBO J *20*, 924-932.

Kuduvalli, P. N., Mitra, R., and Craig, N. L. (2005). Site-specific Tn7 transposition into the human genome. Nucleic Acids Res *33*, 857-863.

Lawley, T., Wilkins, B., and Frost, L. (2004). Bacterial conjugation in Gram-negative bacteria. Plasmid Biology, 203–226.

Lichtenstein, C., and Brenner, S. (1981). Site-specific properties of Tn7 transposition into the *E. coli* chromosome. Mol Gen Genet *183*, 380-387.

Lichtenstein, C., and Brenner, S. (1982). Unique insertion site of Tn7 in *E. coli* chromosome. Nature *297*, 601-603.

Lu, F., and Craig, N. L. (2000). Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. Embo J *19*, 3446-3457.

May, E. W., and Craig, N. L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science *272*, 401-404.

McKenzie, G. J., and Craig, N. L. (2006). Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. BMC Microbiol *6*, 39.

McKown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988). Sequence requirements of *Escherichia coli attTn7*, a specific site of transposon Tn7 insertion. J Bacteriol *170*, 352-358.

McKown, R. L., Waddell, C. S., Arciszewska, L. A., and Craig, N. L. (1987). Identification of a transposon Tn7-dependent DNA-binding activity that recognizes the ends of Tn7. Proc Natl Acad Sci USA *84*, 7807-7811.

Mosig, G. (1998). Recombination and recombination-dependent DNA replication in bacteriophage T4. Annu Rev Genet *32*, 379-413.

Oppon, J. C., Sarnovsky, R. J., Craig, N. L., and Rawlings, D. E. (1998). A Tn7-like transposon is present in the *glmUS* region of the obligately hemoautolithotrophic bacterium *Thiobacillus ferrooxidans*. J Bacteriology *180*, 3007-3012.

Orman, B. E., Pineiro, S. A., Arduino, S., Galas, M., Melano, R., Caffer, M. I., Sordelli, D. O., and Centron, D. (2002). Evolution of multiresistance in nontyphoid salmonella serovars from 1984 to 1998 in Argentina. Antimicrob Agents Chemother *46*, 3963-3970.

Osborn, A. M., and Boltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. Plasmid *48*, 202-212.

Parks, A. R., and Peters, J. E. (submitted). Tn7 elements: Engendering diversity from chromosomes to episomes. Plasmid.

Parks, A. R., and Peters, J. E. (2007). Transposon Tn7 is widespread in diverse bacteria and forms genomic islands. J Bacteriol *189*, 2170-2173.

Peters, J. E., and Craig, N. L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol Cell *6*, 573-582.

Peters, J. E., and Craig, N. L. (2001a). Tn7 recognizes target structures associated with DNA replication using the DNA binding protein TnsE. Genes & Dev *15*, 737-747.

Peters, J. E., and Craig, N. L. (2001b). Tn7: smarter than we thought. Nature Reviews/Molecular Cell Biology *2*, 806-814.

Ramirez, M. S., Quiroga, C., and Centron, D. (2005a). Novel rearrangement of a class 2 integron in two non-epidemiologically related isolates of Acinetobacter baumannii. Antimicrob Agents Chemother *49*, 5179-5181.

Ramirez, M. S., Vargas, L. J., Cagnoni, V., Tokumoto, M., and Centron, D. (2005b). Class 2 integron with a novel cassette array in a Burkholderia cenocepacia isolate. Antimicrob Agents Chemother *49*, 4418-4420.

Rao, J. E., Miller, P. S., and Craig, N. L. (2000). Recognition of triple-helical DNA structures by transposon Tn7. Proc Natl Acad Sci USA *97*, 3936-3941.

Ronning, D. R., Li, Y., Perez, Z. N., Ross, P. D., Hickman, A. B., Craig, N. L., and Dyda, F. (2004). The carboxy-terminal portion of TnsC activates the Tn7 transposase through a specific interaction with TnsA. Embo J *23*, 2972-2981.

Sarnovsky, R., May, E. W., and Craig, N. L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J *15*, 6348-6361.

Seringhaus, M., Kumar, A., Hartigan, J., Snyder, M., and Gerstein, M. (2006). Genomic analysis of insertion behavior and target specificity of mini-Tn7 and Tn3 transposons in Saccharomyces cerevisiae. Nucleic Acids Res *34*, e57.

Sharpe, P., and Craig, N. L. (1998). Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. EMBO J *17*, 5822-5831.

Sherratt, D. J. (2003). Bacterial chromosome dynamics. Science *301*, 780-785.

Shi, Q., Parks, A.R., Potter, B.D., Safir, I.J., Luo, Y., Forster, B.M., Peters, J.E. (2008a). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in E. coli. Genetics *In Press*.

Shi, Q. P., AR; Potter, BD; Safir, IJ; Luo, Y; Forster, BM; Peters, JE (2008b). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in E. coli. Genetics *In Press*.

Skelding, Z., Queen-Baker, J., and Craig, N. L. (2003). Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. Embo J *22*, 5904-5917.

Skelding, Z., Sarnovsky, R., and Craig, N. L. (2002). Formation of a nucleoprotein complex containing Tn7 and its target DNA regulates transposition initiation. Embo J *21*, 3494-3504.

Smith, G. (1983). General recombination. In Lambda II (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), pp. 175–209.

Stellwagen, A., and Craig, N. L. (1997a). Avoiding Self: Two Tn7-encoded proteins mediate target immunity in Tn7 transposition. EMBO J *16*, 6823-6834.

Stellwagen, A., and Craig, N. L. (1997b). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein which activates the bacterial transposon Tn7. Genetics *145*, 573-585.

Stellwagen, A., and Craig, N. L. (1998). Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. TIBS *23*, 486-490.

Tang, Y., Lichtenstein, C., and Cotterill, S. (1991). Purification and characterization of the TnsB protein of Tn7, a transposition protein that binds to the ends of Tn7. Nucl Acids Res *19*, 3395-3402.

Waddell, C. S., and Craig, N. L. (1988). Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes & Dev *2*, 137-149.

Wang, J. D., Berkmen, M. B., and Grossman, A. D. (2007). Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis. Proc Natl Acad Sci U S A *104*, 5608-5613.

Weigel, C., and Seitz, H. (2006). Bacteriophage replication modules. FEMS Microbiol Rev *30*, 321-381.

Wolkow, C. A., DeBoy, R. T., and Craig, N. L. (1996). Conjugating plasmids are preferred targets for Tn7. Genes & Dev *10*, 2145-2157.

CHAPTER 2. BIOINFORMATIC ANALYSIS OF Tn7-LIKE ELEMENTS SUGGESTS MECHANISMS OF TRANSPOSON AND HOST GENOME EVOLUTION[2]

2.1. Summary

The bacterial transposon Tn7 maintains two distinct lifestyles, one in horizontally transferred DNA and the other in bacterial chromosomes. Access to these two DNA pools is mediated by two separate target selection pathways. In this chapter, I will discuss how the molecular mechanisms of Tn7-like elements contribute to their diversification and how they affect the evolution of their host genomes. The analysis of over 50 Tn7-like elements provides insight into the evolution of these elements. In addition to the genes required for transposition, Tn7-like elements transport a wide variety of genes that contribute to the success of diverse organisms. I propose that by decisively moving between mobile and stationary DNA pools, Tn7-like elements accumulate a broad range of genetic material, providing a selective advantage for diverse host bacteria.

2.2. Introduction

With the advent of massively parallel high throughput sequencing, many researchers are turning to genomic approaches for answering a wide

[2] This chapter has been submitted as part of a review for the journal Plasmid Parks, A. R., and Peters, J. E. (submitted). Tn7 elements: Engendering diversity from chromosomes to episomes. Plasmid..

variety of questions. As a result, many new microbial genomes have been sequenced and placed in publicly accessible databases for use of the scientific community. These genomes contain immense amounts of information and are an incredible resource. In this chapter, I describe the use of basic bioinformatic tools to search these databases for information regarding Tn7 and Tn7-like elements and to analyze these transposons within the context of a given genome. I also compare the various components of Tn7-like elements to those of Tn7, and use these comparisons to generate hypotheses for further research.

2.3. Results

Phylogeny and diversity of transposition genes

Tn7 and its related elements have been found in a wide variety of bacteria from diverse environments (Parks and Peters, 2007). They can be found in, *Firmicutes*, *Cyanobacteria*, *Bacteroidetes*, in nearly all classes of *Proteobacteria,* and in the more deeply branching *Chloroflexi* phylum (Figure 2.1, Table 2.1). The ecological niches occupied by Tn7 hosts are quite diverse. These organisms can be found in deep sea hydrothermal vents (*Idiomarina loihiensis, Shewanella loihica*), in low pH acid mine drainage (*Acidithiobacillus ferrooxidans*) and high pH soda lakes (*Natranaerobius thermophilus JW/NM-WN-LF*), in the surface waters of the ocean (oceanic environmental metagenome libraries), in spoiled food (*Bacillus cereus*), in soil

TABLE 2.1. Tn7-like elements

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Escherichia coli (Gammaproteobacteria), **NC_002525**	plasmid, *attTn7*	TnsA	NP_065320.1
		TnsB	NP_065319.1
	Position of ends: R-24497, L-10431	TnsC	NP_065318.1
		TnsD	NP_065317.1
		TnsE	NP_065316
Acinetobacter baumannii (SDF) (Gammaproteobacteria), **NC_010400**	*attTn7*	TnsA	CAP02825.1
		TnsB	CAP02824.1
	Position of ends: R-3310924, L-3276082	TnsC	CAP02823.1
		TnsD	CAP02822.1
		TnsE	CAP02821.1
Aeromonas salmonicida subsp. salmonicida A449 (Gammaproteobacteria), **NC_009348**	*attTn7*	TnsA	YP_001144004.1
		TnsB	YP_001144003.1
	Position of ends: R-4646261, L-4625451	TnsC	YP_001144002.1
		TnsD	Pseudo (gi|145297124)
		TnsE	YP_001143999
Shewanella loihica PV-4 (Gammaproteobacteria), **NC_009092**	*attTn7*	TnsA	YP_001095953.1
		TnsB	YP_001095952.1
	Position of ends: R-4568999, L-4548024	TnsC	YP_001095951.1
		TnsD	YP_001095950.1
		TnsE	YP_001095949
Acinetobacter baumannii ATCC 17978 (Gammaproteobacteria), **NC_009085**	non-*attTn7*	TnsA	YP_001085569.1
		TnsB	YP_001085570.1
	Position of ends: R-2944622, L-2989589	TnsC	YP_001085571.1
		TnsD	YP_001085572.1, YP_001085574.1
		TnsE	YP_001085575

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Moritella sp. PE36 (Gammaproteobacteria), **NZ_ABCQ01000022.1**	*attTn7*	TnsA	ZP_01898871.1
		TnsB	ZP_01898872.1
	Position of ends: R-18747, L-51006	TnsC	ZP_01898873.1
		TnsD	ZP_01898874.1
		TnsE	ZP_01898875
Idiomarina loihiensis L2TR (Gammaproteobacteria), **NC_006512**	*attTn7*	TnsA	YP_156994.1
		TnsB	YP_156993.1
	Position of ends: R-2816674, L-2800435	TnsC	YP_156992.1
		TnsD	YP_156991.1
		TnsE	YP_156990
Photobacterium sp. SKA34 (Gammaproteobacteria), **NZ_AAOU01000022.1**	*attTn7*	TnsA	ZP_01161191.1
		TnsB	ZP_01161192.1
	Position of ends: ND	TnsC	ZP_01161193.1
		TnsD	ZP_01161194.1
Shewanella baltica OS155 (Gammaproteobacteria), **NC_009052**	*attTn7,* non-*attTn7*	TnsA	YP_001052690.1, YP_001052657.1
		TnsB	YP_001052689.1, YP_001052656.1
	Position of ends: R-5103538, L-5079635, ND	TnsC	YP_001052688.1, YP_001052652.1
		TnsD	YP_001052687.1, YP_001052651.1
		TnsE	YP_001052686

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Shewanella putrefaciens CN-32 (Gammaproteobacteria), **NC_009438**	*attTn7*	TnsA	YP_001185451.1, pseudo (gi|146291111)
		TnsB	YP_001185450.1, YP_001185440.1
	Position of ends: R-4596246, L-4575371, R-4575365, L-456489	TnsC	YP_001185449.1, YP_001185439.1
		TnsD	YP_001185448.1, YP_001185436.1
		TnsE	YP_001185447, YP_001185435
Shewanella putrefaciens 200 (Gammaproteobacteria), **NZ_AAWY01000023.1**	*attTn7*	TnsA	ZP_01706553.1
		TnsB	ZP_01706552.1
	Position of ends: R-2083 (contig127), L-3652 (contig 126)	TnsC	ZP_01706551.1
		TnsD	ZP_01706550.1
		TnsE	ZP_01706608
Hahella chejuensis KCTC 2396 (Gammaproteobacteria), **NC_007645**	*attTn7*	TnsA	YP_438104.1
		TnsB	YP_438103.1
	Position of ends: R-7193132, L-7147046	TnsC	YP_438102.1
		TnsD	YP_438101.1
		TnsE	YP_438100

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Acidithiobacillus ferrooxidansATCC 23270 (Gammaproteobacteria), **AF032884.1, also CMR-TIGR:243159**	*attTn7,* non-*attTn7*	TnsA	AAC21667.1, AFE1880, AFE3133
		TnsB	AAC21666.1, AFE1879, AFE3134
		TnsC	AAC21665.1, AFE1878, AFE3133
	Position of ends: R-1762162, L-1731108, R-2900824, L-2939832	TnsD	AAC21664.1, AFE1877, AFE3132
		TnsE	AAC21663, AFE1876, AFE3131
Rhodobacterales bacterium HTCC2255 (Alphaproteobacteria), **NZ_AATR01000005.1**	*attTn7*	TnsA	ZP_01448252.1
		TnsB	ZP_01448253.1
	possition of ends: ND	TnsC	ZP_01448254.1
		TnsD	ZP_01448255.1
Anabaena variabilis ATCC 29413 (Cyanobacteria), **NC_007413**	*attTn7*	TnsA	YP_323988.1
		TnsB	YP_323989.1
	Possition of ends: R-4324033, L-4349748	TnsC	YP_323990.1
		TnsD	YP_323991.1, YP_323997.1
Pelobacter carbinolicus DSM 2380 (Deltaproteobacteria), **NC_007498**	*attTn7*	TnsA	YP_358337.1
		TnsB	YP_358336.1
	Position of ends: ND	TnsC	YP_358335.1
		TnsD	YP_358327.1
		TnsE	YP_358326

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Vibrionales bacterium SWAT-3 (Gammaproteobacteria), **NZ_AAZW01000012.1**, **NZ_AAZW01000014**, **NZ_AAZW01000044**, **NZ_AAZW01000019**	*attTn7,* non-*attTn7*	TnsA	ZP_01813510.1, ZP_01815428.1, ZP_01813648.1, ZP_01814066.1
	Possition of ends: R-116600, L-105729, R-38142, L-27364, R-38129, L-27352, R-39675	TnsB	ZP_01813509.1, ZP_01815427.1, ZP_01813647.1, ZP_01814065.1
		TnsC	ZP_01813508.1, ZP_01815426.1, ZP_01813646.1, ZP_01814064.1
		TnsD	ZP_01813507.1, ZP_01815424.1,ZP_01815425.1, ZP_01813645.1, ZP_01814063.1
		TnsE	ZP_01813506, ZP_01815423.1, ZP_01813644.1, ZP_01814062.1
Marinobacter sp. ELB17 (Gammaproteobacteria), **NZ_AAXY01000016.1**	*attTn7*	TnsA	ZP_01738582.1
		TnsB	ZP_01738583.1
	Position of ends: ND	TnsC	ZP_01738584.1
		TnsD	ZP_01738598.1
		TnsE	ZP_01738599

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Bacillus cereus ATCC 10987 (Firmicutes), **NC_003909**	*attTn7*	TnsA	NP_976510.1, NP_976502.1
		TnsB	NP_976511.1, NP_976503.1
	Possition of ends: R-159705, L-187174, R187178, L-198135	TnsC	NP_976512.1, NP_976504.1
		TnsD	NP_976513.1, NP_976505.1, AAS39114.1
		TnsE	NP_976514
Staphylococcus sp. 693-2 (Firmicutes), **NC_009130**	*plasmid, pLEW6932*	TnsA	YP_001096301.1
		TnsB	YP_001096300.1
	Possition of ends: R-30744, L-17105	TnsC	YP_001096299.1
		TnsD	YP_001096298.1
		TnsE	YP_001096297
Herpetosiphon aurantiacus ATCC 23779 (Chloroflexi), **NC_009972**	*attTn7*	TnsA	YP_001546711.1
		TnsB	pseudo (gi|159896533)
	Possition of ends: ND	TnsC	YP_001546712.1
		TnsD	YP_001546713.1
Clostridium thermocellum ATCC 27405 (Firmicutes), **NC_009012**	non-*attTn7*	TnsA	YP_001037542.1, YP_001038290.1
		TnsB	YP_001037541.1, YP_001038287.1
	Possition of ends: ND	TnsC	YP_001037540.1, YP_001038286.1
		TnsD	YP_001037539.1, YP_001038285.1

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Burkholderia phymatum STM815 (Betaproteobacteria), **NC_010623**	non-*attTn7*	TnsA	ZP_01505407.1, ZP_01500254.1
		TnsB	ZP_01505406.1, ZP_01500253.1
	Possition of ends: ND	TnsC	ZP_01505405.1, ZP_01500252.1
		TnsD	ZP_01505404.1, ZP_01505428.1
Bacillus cereus H3081.97 (Firmicutes), **NZ_ABDL01000038.1**	*attTn7*	TnsA	ZP_02598971.1
		TnsB	ZP_02598972.1
	Possition of ends: R-24658, L-36935	TnsC	ZP_02598973.1
		TnsD	YP_323997.1
		TnsE	ZP_02598974
Pseudomonas fluorescens PfO-1 (Gammaproteobacteria), **NC_007492**	*attTn7*	TnsA	YP_351449.1
		TnsB	YP_351448.1
	Possition of ends: ND	TnsC	YP_351447.1
		TnsD	YP_351445.1
Herminiimonas arsenicoxydans (Betaproteobacteria), **NC_009138**	non-*attTn7*	TnsA	YP_001100471.1
		TnsB	YP_001100472.1
	Possition of ends: ND	TnsC	YP_001100473.1
		TnsD	YP_001100474.1
Nostoc punctiforme PCC 73102 (Cyanobacteria), **NZ_AAAY02000030.1**	non-*attTn7*	TnsA	ZP_00109788.1
		TnsB	ZP_00109790.1
	Possition of ends: ND	TnsC	ZP_00109791.2
		TnsD	ZP_00109794.2

TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Clostridium butyricum 5521 (Firmicutes), **NZ_ABDT01000003.2**	*attTn7*	TnsA	ZP_02623272.1, ZP_02623267.1, ZP_02623280.1
		TnsB	ZP_02623273.1, ZP_02623268.1, ZP_02623281.1
	Possition of ends: R-70986, L-77464, R-77464, L-92891, R-92896, L-109130, R-109135	TnsC	ZP_02623274.1, ZP_02623282.1,
		TnsD	ZP_02623275.1, ZP_02623276.1, ZP_02623283.1
		TnsE	ZP_02623284.1
Escherichia coli APEC O1 (Gammaproteobacteria), **NC_009838**	plasmid, pAPEC-01-R	TnsA	ABF67781.1
		TnsB	ABF67782.1
	Possition of ends: R-109006, L-142480	TnsC	ABF67783.1
		TnsD	ABF67784.1, ABF67785.1
Serratia marcescens (Gammaproteobacteria), **NC_005211**	plasmid, R478	TnsA	NP_941203.1
		TnsB	NP_941204.1
	Possition of ends: R-114733, L-147141	TnsC	NP_941205.1
		TnsD	NP_941206.1, CAE51663.1
Ralstonia metallidurans CH34 (Betaproteobacteria), **NC_007974**	non-*attTn7*	TnsA	YP_587587.1
		TnsB	YP_587586.1
	Possition of ends: ND	TnsC	YP_587585.1
		TnsD	YP_585913.1

TABLE 2.1 (Continued)

Organism, and accession number	Genomic Context	Protein accession number	
Natranaerobius thermophilus JW/NM-WN-LF (Firmicutes), **NZ_ABKR01000020.1**	*attTn7*	TnsA	ZP_02853428.1
		TnsB	ZP_02853427.1
	Possition of ends: ND	TnsC	ZP_02853426.1
		TnsD	ZP_02853425
		TnsE	ZP_02853424
Photobacterium sp. SKA34 (Gammaproteobacteria), **NZ_AAOU01000013.1**	non-*attTn7*	TnsA	ZP_01160363.1
		TnsB	ZP_01160362.1
	Possition of ends: ND	TnsC	ZP_01160359.1
		TnsD	ZP_01160358.1
Herminiimonas arsenicoxydans (Betaproteobacteria), **NC_009138**	non-*attTn7*	TnsA	YP_001101453.1
		TnsB	YP_001101452.1
	Possition of ends: ND	TnsC	YP_001101451.1
		TnsD	YP_001101450.1
Aeromonas salmonicida subsp. salmonicida A449 (Gammaproteobacteria), **NC_009349**	non-*attTn7*	TnsA	YP_001144065.1
		TnsB	YP_001144064.1
	Possition of ends: ND	TnsC	YP_001144063.1
		TnsD	YP_001144060.1
Janthinobacterium sp. Marseille (Betaproteobacteria), **NC_009659**	non-*attTn7*	TnsA	YP_001355144.1
		TnsB	YP_001355143.1
	Possition of ends: ND	TnsC	YP_001355142.1
		TnsD	YP_001355141.1
Psychromonas ingrahamii 37 (Gammaproteobacteria), **NC_008709**	non-*attTn7*	TnsA	YP_944405.1
		TnsB	YP_944404.1
	Possition of ends: ND	TnsC	YP_944403.1
		TnsD	YP_944401.1
Rhodoferax ferrireducens T118 (Betaproteobacteria), **NC_007908**	non-*attTn7*	TnsA	YP_522441.1
		TnsB	YP_522442.1
	Possition of ends: ND	TnsC	YP_522443.1
		TnsD	YP_522444.1
Vibrio cholerae RC385 (Gammaproteobacteria), **DS265382.1**	non-*attTn7*	TnsA	EDN12385.1
		TnsB	EDN12384.1
	Possition of ends: ND	TnsC	EDN12383.1
		TnsD	EDN12382.1

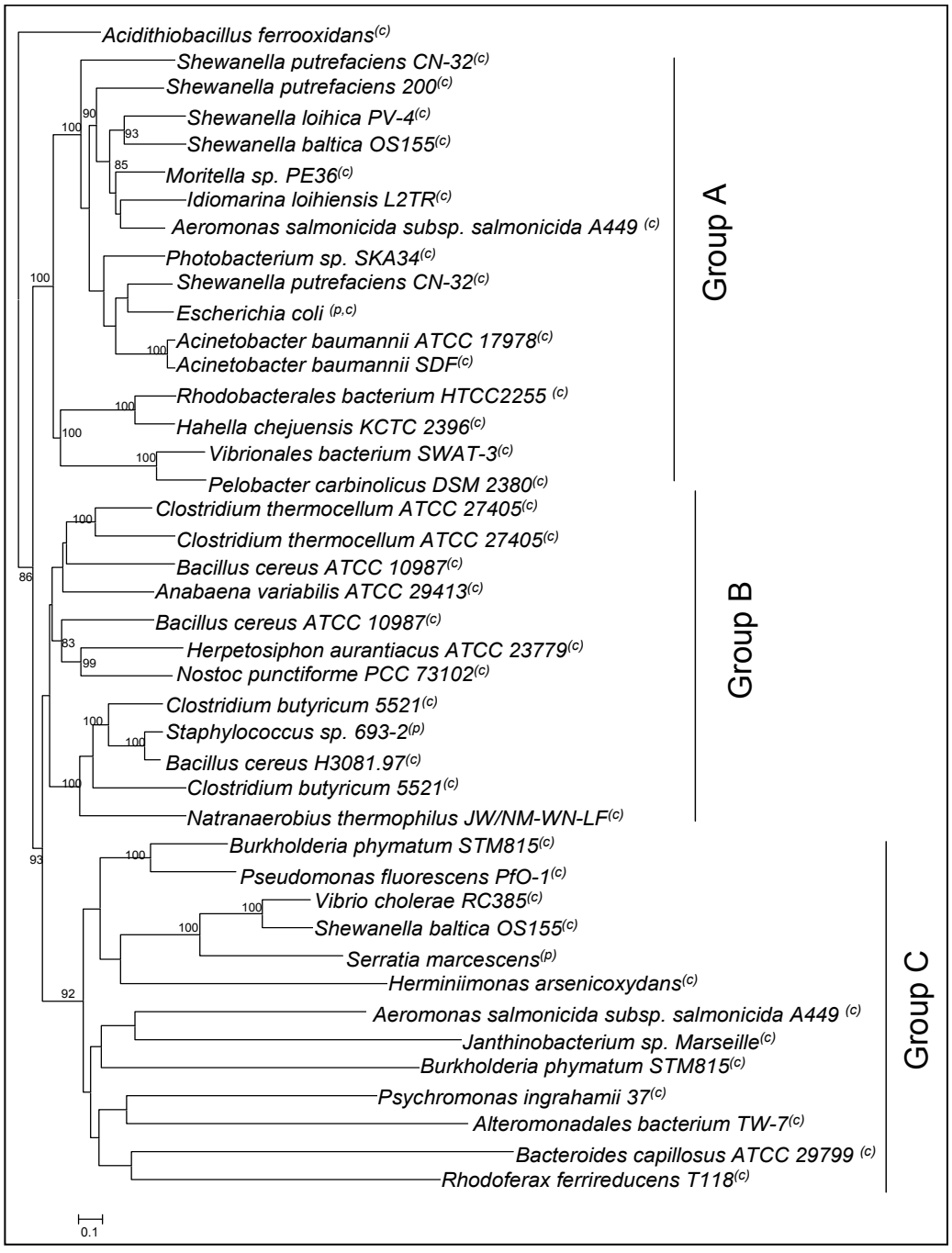
TABLE 2.1 (Continued)

Organism, and accession number	**Genomic Context**	**Protein accession number**	
Alteromonadales bacterium TW-7 (Gammaproteobacteria), **NZ_AAVS01000026.1**	non-*attTn7*	TnsA	ZP_01614175.1
		TnsB	ZP_01614176.1
	Possition of ends: ND	TnsC	ZP_01614177.1
		TnsD	ZP_01614179.1
Burkholderia cenocepacia MC0-3 (Betaproteobacteria), **NZ_AAVA01000005.1**	non-*attTn7*	TnsA	ZP_01562008.1
		TnsB	ZP_01562105.1
	Possition of ends: ND	TnsC	ZP_01562007.1
		TnsD	ZP_01562104.1
Bacteroides capillosus ATCC 29799 (Bacteroidetes), **NZ_AAXG02000042.1**	non-*attTn7*	TnsA	ZP_02038365.1
		TnsB	ZP_02038364.1
	Possition of ends: ND	TnsC	ZP_02038363.1
		TnsD	ZP_02038362.1

samples (soil environmental metagenomic libraries) and in clinical settings around the world (*Escherichia coli, Helicobacter pylori, Burkholderia cenocepacia, Pseudomonas aeruginosa, Shigella sonnei, Clostridium butyricum* 5521, and others) to name just a few. The G+C content of *tns* genes from each of the elements typically mirrors that of the host genome in which they reside and ranges from 26% (in *Clostridium butyricum* 5521) to 58% (in *Acidithiobacillus ferrooxidans*), suggesting that these elements have been maintained within similar organisms for some time. However, in some cases, such as in *Hahella chejuensis*, there are significant differences in G+C content. Differences in G+C content can be taken as evidence of recent arrival of the Tn7-like element (Lawrence and Ochman, 1997).

Phylogenetic analysis of the transposition genes of Tn7-like elements reveals three distinct lineages that I have designated Groups A, B, and C (Figure 2.1). Group A is comprised of elements found within *Proteobacteria*, and includes the original Tn7 from *E. coli*. Group B elements can be found in *Firmicultes*, *Cyanobacteria*, and *Chloroflexi*. Although found in a *Gammaproteobacterium*, the elements found in *Acidithiobacillus ferrooxidans* are most similar to Group B. Group C elements are very different from those in Groups A and B in that none of them seem to contain a *tnsE* gene. Surprisingly, none are found within the traditional *attTn7* locus even though they all contain *tnsD* homologs. Some of the Group C elements appear to be near the *attTn7* locus, but the orientation of *tns* genes and their proximity to the *glmS* gene is not consistent with standard Group A and B Tn7-like elements. It therefore seems likely that the TnsD proteins in the Group C

Figure 2.1. Phylogenetic analysis of Tn7-like elements reveals three distinct lineages of transposons. The predicted amino acid sequences of TnsABCD from each element was concatenated and aligned using the Jalview software package and the ClustalW algorithm (see Table 2.1 for accession numbers) (Chenna et al., 2003; Clamp et al., 2004). Protein distances and Neighbor Joining trees were constructed using the Phylip program (Felsenstein, 1993). The tree was rooted on *Acidithiobacillus ferroxidans*. Bootstrap analysis was also carried out using 100 iterations, and values greater than 80% are given at appropriate branch points. The phylogenetic tree was drawn using the MEGA4 software package (Tamura et al., 2007). “C” indicates chromosomally localized, and “P” indicates that the element is found on a plasmid.


Acidithiobacillus ferrooxidans(c)
Shewanella putrefaciens CN-32(c)
Shewanella putrefaciens 200(c)
Shewanella loihica PV-4(c)
Shewanella baltica OS155(c)
Moritella sp. PE36(c)
Idiomarina loihiensis L2TR(c)
Aeromonas salmonicida subsp. salmonicida A449 (c)
Photobacterium sp. SKA34(c)
Shewanella putrefaciens CN-32(c)
Escherichia coli (p,c)
Acinetobacter baumannii ATCC 17978(c)
Acinetobacter baumannii SDF(c)
Rhodobacterales bacterium HTCC2255 (c)
Hahella chejuensis KCTC 2396(c)
Vibrionales bacterium SWAT-3(c)
Pelobacter carbinolicus DSM 2380(c)
Clostridium thermocellum ATCC 27405(c)
Clostridium thermocellum ATCC 27405(c)
Bacillus cereus ATCC 10987(c)
Anabaena variabilis ATCC 29413(c)
Bacillus cereus ATCC 10987(c)
Herpetosiphon aurantiacus ATCC 23779(c)
Nostoc punctiforme PCC 73102(c)
Clostridium butyricum 5521(c)
Staphylococcus sp. 693-2(p)
Bacillus cereus H3081.97(c)
Clostridium butyricum 5521(c)
Natranaerobius thermophilus JW/NM-WN-LF(c)
Burkholderia phymatum STM815(c)
Pseudomonas fluorescens PfO-1(c)
Vibrio cholerae RC385(c)
Shewanella baltica OS155(c)
Serratia marcescens(p)
Herminiimonas arsenicoxydans(c)
Aeromonas salmonicida subsp. salmonicida A449 (c)
Janthinobacterium sp. Marseille(c)
Burkholderia phymatum STM815(c)
Psychromonas ingrahamii 37(c)
Alteromonadales bacterium TW-7(c)
Bacteroides capillosus ATCC 29799 (c)
Rhodoferax ferrireducens T118(c)
0.1
Group A
Group B
Group C

elements actually catalyze a more random form of transposition. These elements may represent a more primitive form of Tn7. There are at least two examples in which a Group C element inhabits the same organism as a Group A element. Further study of Group C elements and their relationship to Groups A and B may provide some interesting clues regarding the evolution of Tn7-like transposons. We have yet to find any Tn7-like elements in the Archaea, even though genetic exchange is known to occur between Bacteria and Archaea (Nelson et al., 1999).

Experiments have shown that the chromosome targeting pathway of Tn7 is extremely robust (Choi et al., 2005; McKenzie and Craig, 2006). Indeed, this pathway is capable of directing transposition into a wide variety of host genomes in the laboratory including into human "*attTn7*" loci associated with *gfpt*-1 and *gfpt*-2, the *glmS* homologs found in our genome (Kuduvalli et al., 2005). The preponderance of Tn7 elements that have been discovered to date have been found within the presumed *attTn7* locus of organisms with draft or completely sequenced genomes. This likely reflects a database bias in which there are more genomic DNA sequences present than viral or plasmid sequences. Alternatively, plasmid and bacteriophage borne Tn7s may be far more transient since excision of the transposon may destroy the vector by the resulting DNA double-strand break. There are, however, some examples, both in gram (+) and in gram (-) bacteria, of plasmids containing Tn7-like elements. No Tn7 homologs have yet been found in bacteriophages outside of the laboratory, however many chromosomal Tn7-like elements either contain genes associated with bacteriophages or are inserted in the vicinity of phage-associated genes (Parks and Peters, 2007) (Figure 2.2). It seems likely that bacteriophage borne elements will soon be discovered. As

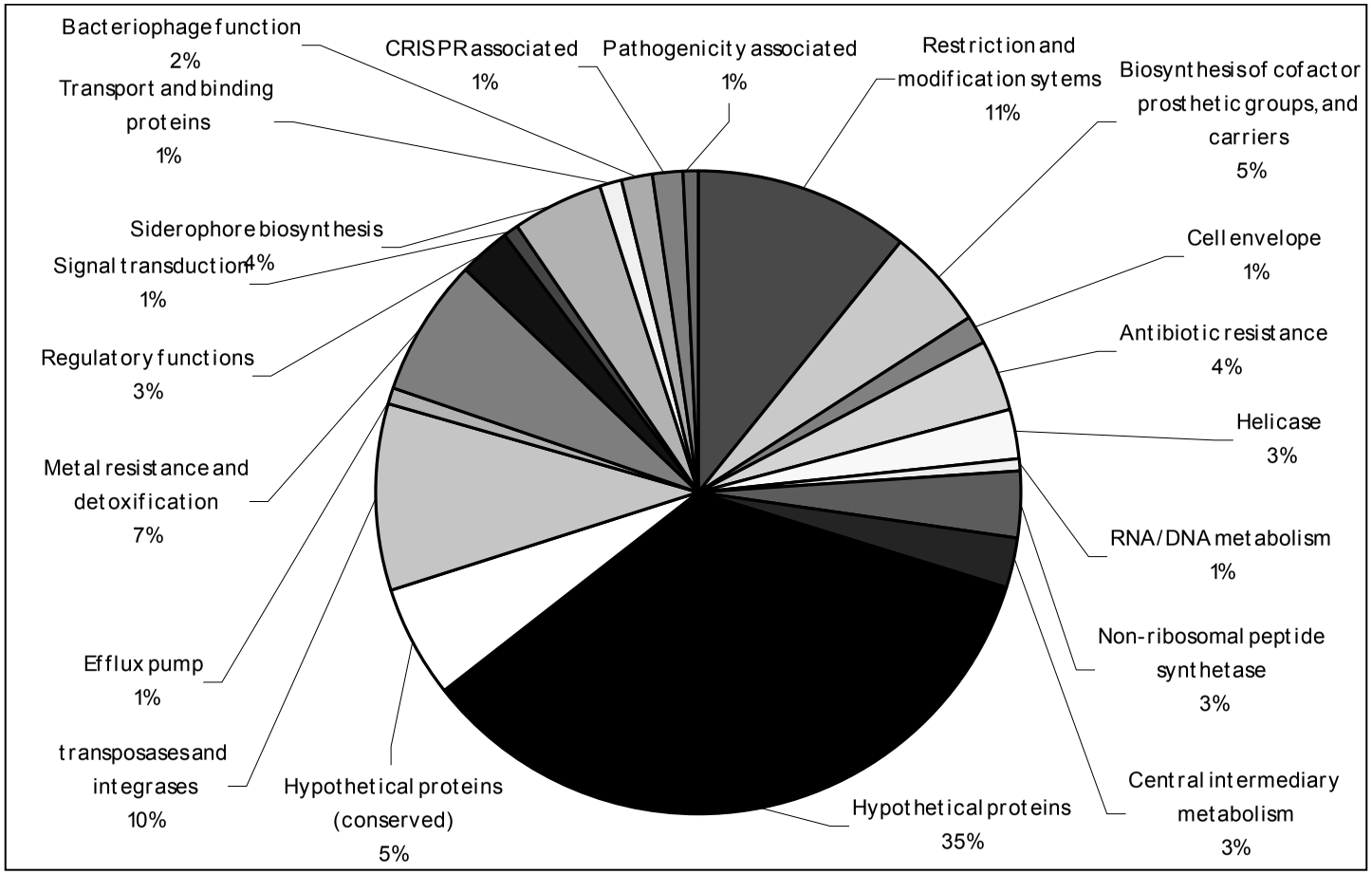
genome and meta-genomic sequencing projects progress, the bestiary of Tn7-like elements continues to mount in population.

Genes transported by Tn7-like elements

In addition to the genes described above that carry out transposition, Tn7 contains a class II integron system that is composed of an integrase gene, *intI2*, along with antibiotic resistance gene cassettes *dhfr*, *sat*, and *aadA,* each flanked by class II integron recombination sequences. The integrase gene is interrupted by an ochre stop codon, which can be found at the same position in all sequenced *intI2* genes (Hansson et al., 2002). Based on its conservation, it has been proposed that the stop codon within the integrase is in fact a regulatory mechanism, not simply a random nonsense mutation, preventing wanton recombination of the integron under normal conditions (Hansson et al., 2002). Elements that appear to be otherwise identical to the original Tn7 contain additional integron cassettes that encode genes for chloramphenicol and kanamycin resistance, among others, suggesting that recombination within the integron region of Tn7 is indeed active (Biskri and Mazel, 2003; Orman et al., 2002; Ramirez et al., 2005a; Ramirez et al., 2005b). Elements that are only slightly divergent from Tn7 contain entirely different integron systems (Barlow and Gobius, 2006; Parks and Peters, 2007).

While the integron cassette system of Tn7 will allow genetic content to vary significantly through the integron-catalyzed shuffling of gene cassettes, other Tn7-like elements also encode a wide variety of genes apart from or instead of integron cassettes. The contents of the various Tn7-like elements can be determined by locating the ends of the elements and analyzing the genes encompassed by the ends and searching for homologs in various gene

Figure 2.2. Tn7-like elements encode diverse arrays of genes with many functions. 280 Genes were selected from elements in which the right and left ends have been determined (Table 2.1). The function of genes found between the ends of the Tn7-like elements were surmised using the BLAST algorithm to search gene databases for related genes with known functions. Genes were then grouped by functional categories of. Functional categories of genes are loosely modeled after CMR-TIGR functional categories (Peterson et al., 2001).


Bacteriophage function
2%
CRISPR associated
1%
Pathogenicity associated
1%
Restriction and
modification sytems
11%
Biosynthesis of cofactor
prosthetic groups, and
carriers
5%
Transport and binding
proteins
1%
Siderophore biosynthesis
4%
Signal transduction
1%
Cell envelope
1%
Antibiotic resistance
4%
Regulatory functions
3%
Helicase
3%
Metal resistance and
detoxification
7%
RNA/DNA metabolism
1%
Non-ribosomal peptide
synthetase
3%
Efflux pump
1%
transposases and
integrases
10%
Hypothetical proteins
(conserved)
5%
Hypothetical proteins
35%
Central intermediary
metabolism
3%

databases (Parks and Peters, 2007). This type of analysis reveals that elements that have nearly identical *tnsABCDE* genes can contain vastly different genetic cargos (Parks and Peters, 2007). There appears to be no unifying theme with regard to the types of genes that are carried by Tn7-like elements. Tn7-like elements contain genes involved in metal resistance and detoxification, putative DNA repair enzymes and polymerases, DNA restriction and modification systems, non-ribosomal peptide synthesis modules, siderophore production genes, efflux pumps, transcription regulators, bacteriophage associated genes, additional transposases and integrases, and many hypothetical genes with unknown functions (Figure 2.2).

DNA restriction and modification systems are an especially common feature in Tn7-like elements. These systems likely serve as host addiction systems in which the host genome is modified by a labile methyltransferase and all DNA that lacks the methylation signature is restricted by a more stable endonuclease protein, preventing loss of the transposon that encodes the methyltransferase (Kobayashi, 2004). These same systems may also help protect host cells from invasion of potentially deleterious foreign DNA by restricting intruding DNAs (i.e. bacteriophages and plasmids) before they are able to establish the methylation signature of the host genome (Kobayashi, 2001). The restriction and modification systems may also contribute to propagation of Tn7-like elements by causing DNA double-strand breaks that must be fixed by host machinery for survival. Tn7 is able to use the TnsE pathway to insert proximal to DSBs (Peters and Craig, 2000; Shi, 2008) and may be able to use this facet of restriction and modification systems to move to new, potentially mobile, DNA molecules when stresses prevent sufficient methyltransferase activity. It is possible that DNAs that are easily fixed by

homologous recombination once restricted, such as multicopy plasmids or lytic bacteriophages, may be activated as targets for transposition once methyltransferase activity reaches a critically low level.

Mechanisms of host protection in general appear to be a common theme in Tn7-like transposons. There is one example of a Tn7-like element within the genome of *Anabaena variabilis ATCC 29413* in which a CRISPR (clustered, regularly interspaced short palindromic repeats) locus can be found. CRISPR regions have recently been shown to constitute a primitive immune system in Archea and Bacteria (Sorek et al., 2008). These systems provide protection from a broad range of invasive bacteriophages, plasmids, and transposons. CRISPR loci have previously been described on plasmids (Godde and Bickerton, 2006), but to our knowledge, this is the first example in which the system has been found within a transposon.

In addition to the known medical importance of Tn7, some organisms that have emerged as pathogenic threats carry Tn7-like elements with especially intriguing genetic contents. A Tn7-like element within the virulent and multiply antibiotic resistant *Acinetobacter baumanii* ATCC 17978 contains a probable multidrug eflux pump, siderophore genes, and non-ribosomal peptide synthesis genes (Vallenet et al., 2008), suggesting that the transposon may contribute significantly to the pathogenicity and persistence of this organism (Iacono et al., 2008). A closely related yet non-virulent strain (*Acinetobacter baumanii SDF*) also contains a highly similar Tn7-like element which lacks the siderophore production and antibiotic resistance genes (Iacono et al., 2008), underscoring the variability amongst Tn7-like elements and their potential importance to host success. A Tn7-like element found in the type E toxin producing *Clostridium butyricum* 5521, a causative agent of

botulism, contains two genes found within a lambdoid prophage of the *Bacillus anthracis* Ames strain (Read et al., 2003). While these genes (an acetyltransferase and a hypothetical gene) are not thought to contribute directly to pathogenicity, they may affect pathogen-host interactions and enhance the organisms ability to survive within a host (Read et al., 2003).

In most Tn7-like elements, the diverse arrays of genes reside in a highly variable region in the left end of the transposon (see Chapter1, Figure 1.1). Maintaining the genes involved in transposition in the right end of the transposon appears to be important for the function of the transposon in some way; the *tns* genes are almost always localized to the extreme right end of the element and with a highly conserved synteny. The integron system of Tn7 from *E. coli* (and others) and the CRISPR system of the Tn7-like element in *Anabaena variabilis* include mechanisms that add DNA sequences to specific regions. There are also unrelated mobile elements that actively target sequences associated integron cassettes (Quiroga et al., 2008). While there may be other yet-to-be-described mechanisms that Tn7-like elements exploit for the addition of genetic material, the variable region may simply provide a "neutral zone" where recombination events can occur without causing harm to the organism or disrupting genes essential for transposition, yet allowing widespread distribution of novel gene combinations.

Formation of genomic Islands within the attTn7 locus

The extreme target-site specificity of the TnsD pathway has some very interesting consequences for bacterial genomes. While target site immunity generally prevents multiple insertions of Tn7 within *attTn7*, many organisms contain multiple Tn7-like elements inserted in tandem within this very

specifically defined DNA locus. These insertions are striking in that while some of the elements appear to be highly divergent, they still insert in the exact same position relative to *glmS*. This is evidenced by the fact that they duplicate exactly the same, or very slightly overlapping, 5 bp sequence upon insertion (Parks and Peters, 2007). These elements accumulate within the *attTn7* locus and lead to the formation of genomic islands. In the most extreme example, found in *Clostridium butyricum* 5521, there are three complete or nearly complete Tn7-like elements. The remnants of a right end from a fourth element also reside adjacent to the left end of the last complete element (Table 2.1).

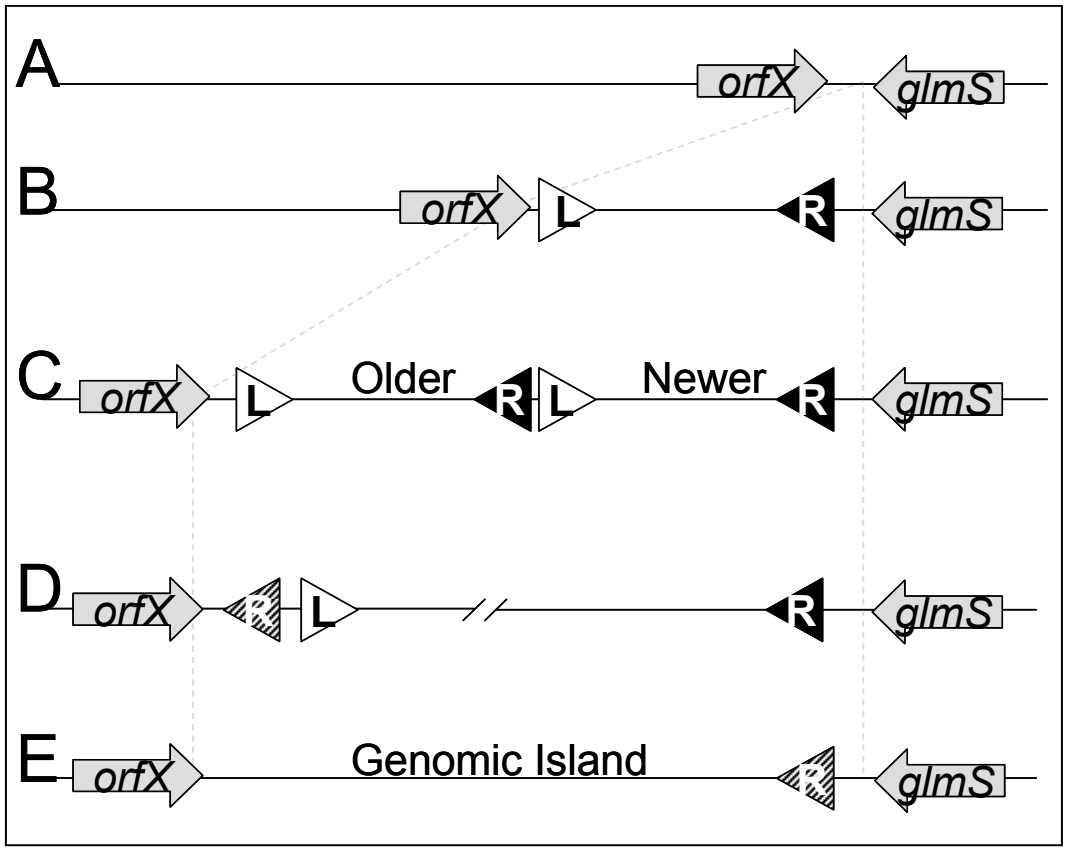
Within the *Shewanella* genus alone, of which many *attTn7* loci have been sequenced, genomic islands can be observed in each of various stages of development (Figure 2.3). *Shewanella oneidensis* MR-1 may be seen as a typical naïve *Shewanella* species with no evidence of Tn7 insertion proximal to the *glmS* gene (Figure 2.3.A.). From *S. oneidensis* and other closely related *Shewanella* species the genes flanking *attTn7* can be identified, indicating that no genes whatsoever are located within *attTn7* locus, making it possible to define the empty *attTn7* with relative certainty (*Shewanella oneidensis* MR-1(AE014299), *S. baltica* OS195 (AATK00000000), *S. sp.* MR-4 (CP000446), *S. sp.* W3-19-1 (AALN01000053), *S. sp.* MR-7 (AALI01000045), *S. sp.* SAR-1 (AACY01051759), and *S. sp.* ANA-3 (AALH01000050)) (Parks and Peters, 2007). *Shewanella putrefaciens* 200 presents an example of an individual insertion (Figure 2.3.B.). *Shewanella putrefaciens* CN-32 is found to possess two Tn7-like insertions in tandem (Figure 2.3.C.) (Parks and Peters, 2007). *Shewanella denitrificans* OS217 (Figure 2.3.E.) maintains horizontally transferred genes that are important for its survival, including denitrification

genes, within the *attTn7* locus yet contains no recognizable Tn7 remnants (Parks and Peters, 2007). It is possible that acquisition of Tn7-encoded genetic material has lead to the specialization that differentiates some *Shewanellae* from one another.

The maximum distance between transposon ends that still allows transposition has not yet been experimentally defined for Tn7. However, the largest single insertion found in the environment is found within *Hahella chejuensis* KCTC 2396, and is 46 Kb in length (Table 2.1). This element is most likely able to transpose in its entirety. The entire length of the element has a significantly lower G+C content than the rest of the genome (47% compared to 53%, respectively), its right and left ends appear to be intact, and all of the genes involved in transposition lack interruptions (Table 2.1). An interesting, but unanswered, question is whether tandem insertions can be mobilized together as a single unit using the outermost right and left ends. The presence of closely related elements within a single organism also presents the possibility of mobilization *in trans*. This process could allow a newly invading Tn7-like element to mobilize previously inactive Tn7 relatives, or complement missing transposon functions in multiple inactivated elements.

As explained in Chapter 1, target site immunity only decreases transposition ~90%. Repeated exposure to Tn7-like elements in the environment could overcome immunity; however, there are also several molecular explanations which could account for the tandem insertion of Tn7-like elements. First, in most cases the transposon ends and proteins are not identical to one another. Nucleotide sequence differences between TnsB binding sites may also prevent heterologous TnsB proteins from recognizing binding sites of a related element (Figure 2.4). Differences in amino acid

Figure 2.3. Schematic depiction of the genesis of genomic islands by Tn7-like transposons. **A.** A representation of a hypothetical *attTn7* site flanked by *glmS* and a hypothetical conserved gene *orfX*. **B**. An individual insertion can be large enough alone to constitute a genomic island. Individual insertions also increase the amount of "neutral" DNA space that may allow other elements to safely insert into the chromosome. **C**. A tandem insertion of another Tn7-like element adjacent to *glmS* expands the *attTn7* locus resulting in newer elements being closer to the *glmS* coding region. **D**. Reductive evolutionary processes, such as recombination, contracts or expands the *attTn7*-associated genomic island and fixes certain genes in place by disabling critical transposon genes and transposon ends. **E**. Further reductive evolution fixes genes with consistent selective pressure and leaves behind only scant evidence of transposon components. Arrows represent open reading frames, while triangles represent transposon ends; (R) for right ends and (L) for left ends. Crosshatched triangles represent degenerate ends that have few TnsB binding sites and poorly conserved sequences.


A
orfX
glmS
B
orfX
L
R
glmS
C
orfX
L
Older
R
L
Newer
R
glmS
D
orfX
R
L
R
glmS
E
orfX
Genomic Island
R
glmS

sequence of either the TnsB or TnsC proteins from different elements may lead to loss of target-sit immunity in a way that maintains the productive interactions required for transposition (see Chapter 1). Some appear to contain three overlapping TnsB binding sites as opposed to the usual four, possibly reducing the local concentration of TnsB enough to have an effect on immunity, but still allowing transposition to occur (Arciszewska and Craig, 1991).

Insertion of multiple Tn7-like elements within the *attTn7* locus may lead to enhanced recombination between divergent Tn7-like elements. As described above, the DSB that is created once an element is mobilized to another location is typically fixed by homologous recombination with the sister chromosome. When multiple Tn7-like elements occupy an attachment site, recombination may occur between neighboring elements creating hybrid elements. Recombination also appears to lead to loss of essential transposition components resulting in the immobilization of transposon-encoded genes within the attachment site (Figure 2.3.E.)(Parks and Peters, 2007).

The ability of Tn7 to form genomic islands may become a useful tool in the laboratory. Tn7 has widely been used to insert genes in single copy into the chromosome of diverse organisms (Choi et al., 2006; Choi et al., 2005; McKenzie and Craig, 2006). Using elements of differing origins may enable investigators to insert various combinations of genes stably into the chromosome with minimal cloning steps.

Analysis of Tn7 ends

When in the *attTn7* locus, Tn7 right-ends are fairly simple to locate, as they typically begin within 50 bp of the end of the *glmS* open reading frame. The 22 bp sequence of the TnsB binding sites can be determined using algorithms that search for tandem repeats, such as the etandem program in the EMBOSS software package, or by visual inspection using genome editing software such as Artemis. The left end of the element can found by inspection of intergenic regions downstream of *tns* gene homologs that are >160 bp. Left ends tend to be more difficult to detect as the binding sites are more degenerate and spaced farther apart. Where necessary, I have employed the Artemis Comparison Tool to help locate the left end of Tn7-like elements. Using this software, whole genomes or contiguous DNA sequences from closely related organisms can be compared, which is sometimes helpful in defining the *attTn7* locus. Using these techniques, I have determined that the exact location of transposon insertion is highly conserved, even when elements within the *attTn7* locus are divergent, such as those found in *Shewanella putrefaciens* CN-32, *Bacillus cereus* ATCC 10987, and *Clostridium butyricum* 5521. These organisms contains two different Tn7-like elements that have duplicated the exact same 5 bp sequence, or overlapping by 1 bp as in the case of *Bacillus cereus* ATCC 10987 (Table 2.1). In each case there are only four or five base-pairs between the terminal ACA from the left end of the newer element and the terminal TGT from the right end of the older element.

Figure 2.4. Alignment of TnsB binding sites in the right end of Tn7-like elements reveals similarities among the elements. A logo displaying the conservation of nucleotides within the TnsB binding sites within group A and B elements displays each base with its size a representation of its conservation, larger letters indicate more highly conserved bases. The logo was constructed by first aligning TnsB binding sites from the right ends of selected elements (*Escherichia coli*, *Shigella sonnei Ss046*, *Pseudomonas aerugenosa*, *Shewanella putrefaciens* CN-32, *Shewanella baltica* OS155, *Shewanella loihica* PV-4, *Idiomarina loihiensis* L2TR, *Acidithiobacillus ferrooxidans* ATCC 23270, *Hahella chejuensis* KCTC 2396, *Pelobacter carbinolicus* DSM 2380, *Bacillus cereus* ATCC 10987, *Staphylococcus sp*.639-2, see Table 2.1 for accession numbers) using the ClustalW program (Chenna et al., 2003), then using the Weblogo algorithm to display homologous sites (Crooks et al., 2004).


bits
2
1
0
Consensus
TGA+AATAAAGTTGATTATACT

The TnsB binding sites from the right ends of the various elements are not identical, however they tend to be A+T rich and asymmetrical (Figure 2.4.).

Analysis of TnsABC homologs

Tn7 has been the subject of intense laboratory research for many years. Many genetic screens have revealed mutations in each of the components of the core machinery that affect target-site selection, regulation and timing of transposition, and the mechanism of transposition. The discovery of multiple Tn7-like elements provides us with the opportunity to examine diverse, yet related, elements and compare them laboratory mutants. This comparison may provide clues to the function of the Tn7-like elements in their natural environment, and provide insight into the evolution of these elements.

TnsA is the endonuclease-like protein that nicks at the 5' phosphates at the ends of the Tn7 element. Interactions between TnsA and TnsC are essential for the regulation of transposition, and certain mutation that affect transposition regulation have been found in TnsA (Lu and Craig, 2000). Some of the naturally occurring TnsA homologs contain the same residues found in genetic screens that dispatch the need for target-site selecting proteins. Some of these mutations, when combined with mutations in TnsB even allow transposition in the absence of the TnsC protein. None of these combinations have been found within a single element, or within a single organism, however the existence of both in naturally occurring elements presents the possibility that combining two transposons within a cell may result in altered regulation of transposition by interaction between the components of the two different core machineries. Some residues found in TnsA homologs might be expected to

decrease or abolish transposition in Tn7 based on the work reported by Lu and Craig,(*E. coli* APEC-01, L70A, Figure 2.5)(Lu and Craig, 2000), illustrating a possible limitation of this type of analysis and caution against conclusions that are drawn from “sequence gazing”. Alternatively, these mutations may point to the possibility of heterologous activation of transposition, by which defective components of an element’s transposition machinery may be complemented by components from a related element.

Homologs of TnsB also contain residues that have been isolated in genetic screens for altered transposon regulation. One example, $TnsB^{M366I}$, appears to be common in many (~8/50) TnsB homologs (Figure 2.6.). This mutation represents a class of TnsB mutation that is stimulated by target-site selection proteins TnsD or TnsE and allows random transposition to occur (see Chapter 3)(Lu and Craig, 2000). This mutation can also be combined with mutations found in TnsA (which also exist in TnsA homologs) to activate transposition in the absence of TnsC. Some of the same amino acid changes that have been identified in TnsB attenuate target-site immunity and can be identified in naturally abundant Tn7-like elements (Figure 2.7)(Skelding et al., 2003; Stellwagen and Craig, 1997). However, the ability to circumvent target-site immunity in these homologs has yet to be determined experimentally. Interestingly, the presumed TnsC-interacting region of TnsB differs among phylogenetic groups (Groups B and C, Figure 2.6.A). The C-terminus of TnsBs is predominantly acidic in one group (Group A, see below), while the same region in other TnsB proteins contain many basic residues (Figure 2.6.A.). In this case differences in TnsB-TnsC interface may define barriers to interaction between groups, but does not explain how closely related elements are capable of tandem insertion within *attTn7*.

Figure 2.5. Alignment of TnsA
TnsA homologs display residues that match those found in genetic screens in the laboratory that affect the regulation of transposition and target-site selection (Lu and Craig, 2000). Sites at which mutations were found in the laboratory are indicated with arrows. Residues that exactly match mutations found in the laboratory are indicated with an asterisk

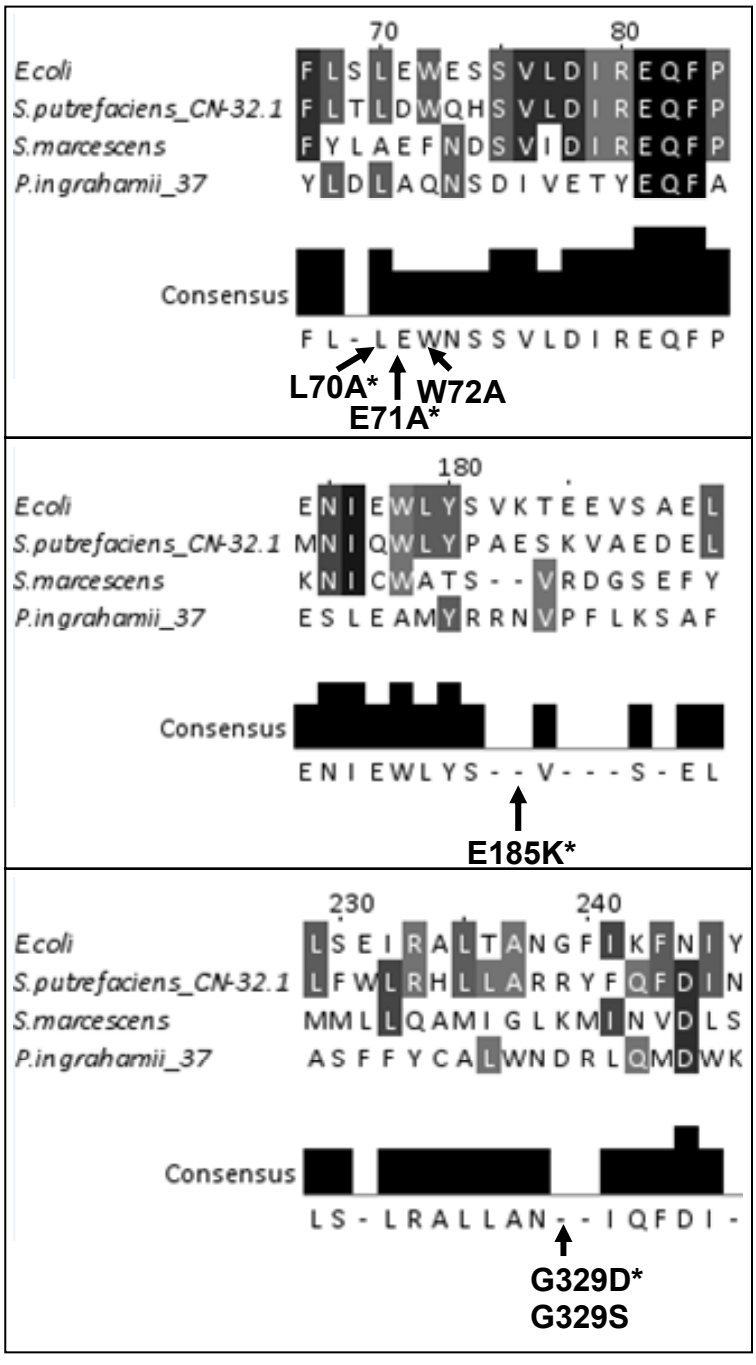
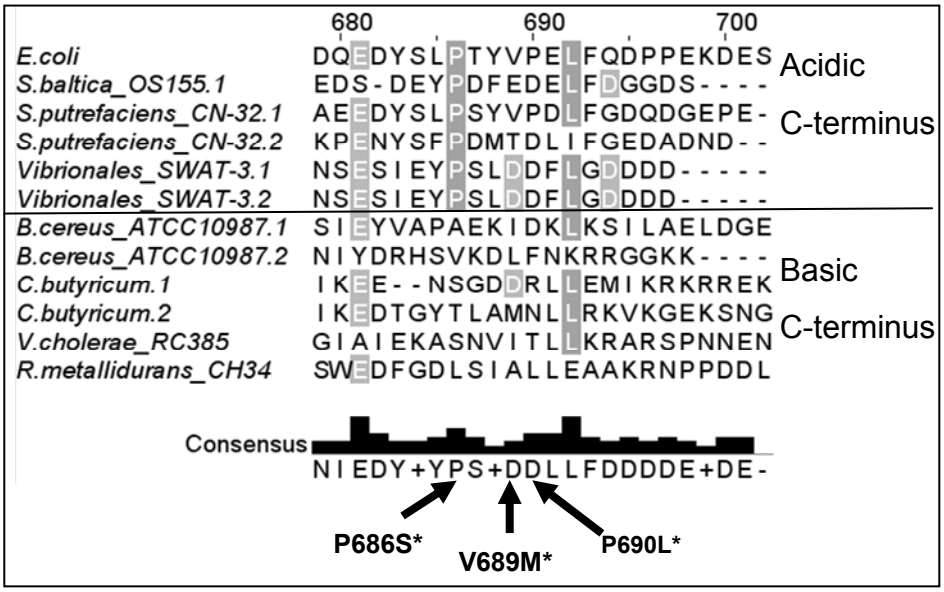

70
80
Ecoli
S.putrefaciens_CN-32.1
S.marcescens
P.ingrahamii_37
Consensus
F L - L E W N S S V L D I R E Q F P
L70A*
E71A*
W72A
180
E185K*
230
240
G329D*
G329S

Figure 2.6. Alignment of TnsB
TnsB proteins reveal possible variations in interaction surfaces between TnsC and TnsB, as well as residues known to result in target-site immunity bypass. A. Sites resulting in target-site immunity bypass are indicated with arrows. Residues that have been found both in TnsB homologs and in mutant screens for loss of target-site immunity are indicated with asterisks. Group A elements contain TnsB proteins with C-termini dominated by acidic residues (aspartate (D) and glutatmate (E)), while Groups B and C contain elements with C-termini residues that are predominantly basic (lysine (K) and arginine (R). TnsC and TnsB homologs from *E. coli* and selected elements (either containing multiple Tn7s within one organism or containing mutations known to affect target-site selection and/or immunity) were aligned using the ClustalW algorithm and edited using the Jalview software package (Chenna et al., 2003; Clamp et al., 2004). B. Some TnsB homologs contain mutations found in laboratory screens that allow transposition in the absence of target-site selecting proteins (Lu and Craig, 2000). There is no experimental evidence for the target selection behavior of these proteins yet, however, these mutations may help explain how some elements end up in locations that are not consistent with TnsD- or TnsE- mediated insertion. Sites where homologs contain the exact residue recovered in mutant screens of TnsB from *E. coli* are marked with an asterisk.

A


680
690
700
E.coli DQEDYSLPTYVPELFQDPPEKDES
S.baltica_OS155.1 EDS-DEYPDFEDELFDGGDS----
S.putrefaciens_CN-32.1 AEEDYSLPSYVPDLFGDQDGEPE-
S.putrefaciens_CN-32.2 KPENYSFPDMTDLIFGEDADND--
Vibrionales_SWAT-3.1 NSESIEYPSLDDFLGDDDD-----
Vibrionales_SWAT-3.2 NSESIEYPSLDDFLGDDDD-----
B.cereus_ATCC10987.1 SIEYVAPAEKIDKLKSILAELDGE
B.cereus_ATCC10987.2 NIYDRHSVKDLFNKRRGGKK----
C.butyricum.1 IKEE--NSGDDRLLEMIKRKRREK
C.butyricum.2 IKEDTGYTLAMNLLRKVKGEKSNG
V.cholerae_RC385 GIAIEKASNVITLLKRARSPNNEN
R.metallidurans_CH34 SWEDFGDLSIALLEAAKRNPPDDL
Acidic
C-terminus
Basic
C-terminus
Consensus
NIEDY+YPS+DDLLFDDDDE+DE-
P686S*
V689M*
P690L*

B

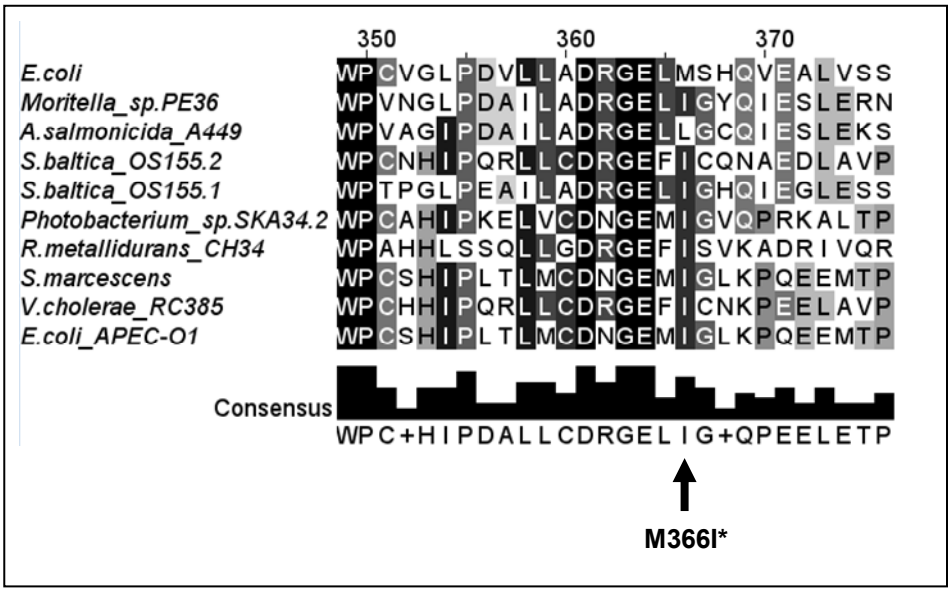
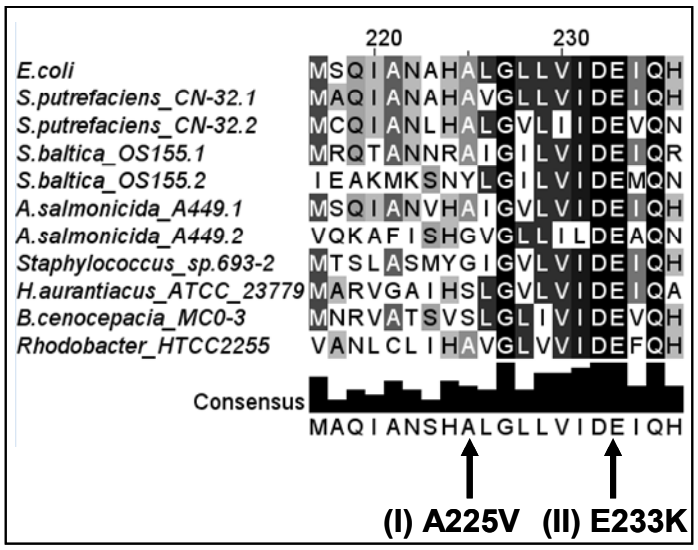
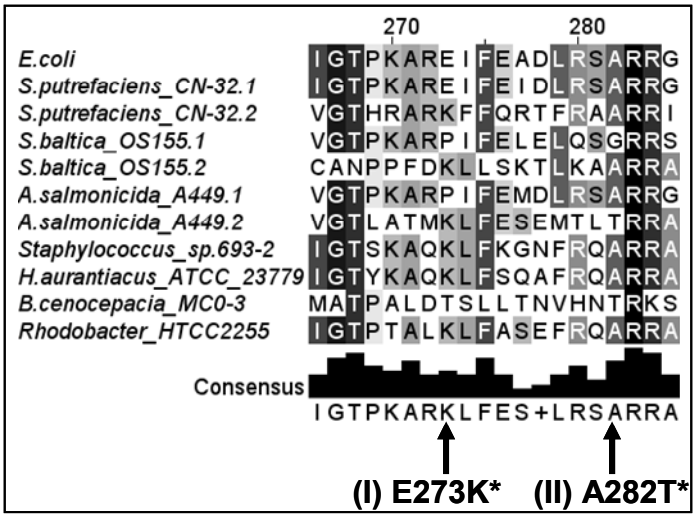
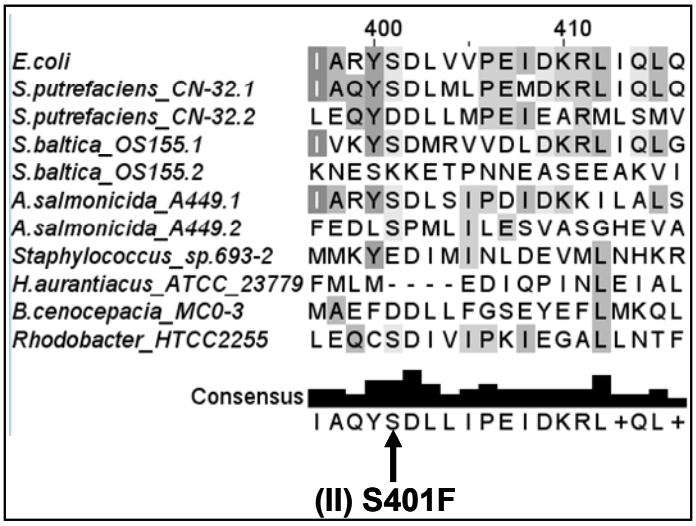

350
360
370
E.coli WPCVGLPDVLLADRGELMSHQVEALVSS
Moritella_sp.PE36 WPVNGLPDAILADRGELIGYQIESLERN
A.salmonicida_A449 WPVAGIPDAILADRGELLGCQIESLEKS
S.baltica_OS155.2 WPCNHIPQRLLCDRGEFICQNAEDLAVP
S.baltica_OS155.1 WPTPGLPEAILADRGELIGHQIEGLESS
Photobacterium_sp.SKA34.2 WPCAHIPKELVCDNGEMIGVQPRKALTP
R.metallidurans_CH34 WPAHHLSSQLLGDRGEFISVKADRIVQR
S.marcescens WPCSHIPLTLMCDNGEMIGLKPQEEMTP
V.cholerae_RC385 WPCHHIPQRLLCDRGEFICNKPEELAVP
E.coli_APEC-O1 WPCSHIPLTLMCDNGEMIGLKPQEEMTP
Consensus
WPC+HIPDALLCDRGELIG+QPEELETP
M366I*

Figure 2.7. Alignment of TnsC
Alignment of selected TnsC homologs reveals possible molecular explanations for insertion into chromosomal sites other than *attTn7*. Sites of Class I (I) and Class II (II) mutations are indicated with arrows. Class I mutations result in random transposition, yet retain sensitivity to target selection signals. Class II mutants are insensitive to target selection signals. Sites where homologs contain the exact residue recovered in mutant screens of TnsC from *E. coli* are marked with an asterisk.


220
230
E.coli MSQIANAHALGLLVIDEIQH
S.putrefaciens_CN-32.1 MAQIANAHAVGLLVIDEIQH
S.putrefaciens_CN-32.2 MCQIANLHALGVLIIDEVQN
S.baltica_OS155.1 MRQTANNRAIGILVIDEIQR
S.baltica_OS155.2 IEAKMKSNYLGILVIDEMQN
A.salmonicida_A449.1 MSQIANVHAIGVLVIDEIQH
A.salmonicida_A449.2 VQKAFISHGVGLLILDEAQN
Staphylococcus_sp.693-2 MTSLASMYGIGVLVIDEIQH
H.aurantiacus_ATCC_23779 MARVGAIHSLGVLVIDEIQA
B.cenocepacia_MC0-3 MNRVATSVSLGLIVIDEVQH
Rhodobacter_HTCC2255 VANLCLIHAVGLVVIDEFQH
Consensus
MAQIANSHALGLLVIDEIQH
(I) A225V
(II) E233K


270
280
E.coli IGTPKAREIFEADLRSARRG
S.putrefaciens_CN-32.1 IGTPKAREIFEIDLRSARRG
S.putrefaciens_CN-32.2 VGTHRARKFFQRTFRAARRI
S.baltica_OS155.1 VGTPKARPIFELELQSGRRS
S.baltica_OS155.2 CANPPFDKLLSKTLKAARRA
A.salmonicida_A449.1 VGTPKARPIFEMDLRSARRG
A.salmonicida_A449.2 VGTLATMKLFESEMTLTRRA
Staphylococcus_sp.693-2 IGTSKAQKLFKGNFRQARRA
H.aurantiacus_ATCC_23779 IGTYKAQKLFSQAFRQARRA
B.cenocepacia_MC0-3 MATPALDTSLLTNVHNTRKS
Rhodobacter_HTCC2255 IGTPTALKLFASEFRQARRA
Consensus
IGTPKARKLFES+LRSARRA
(I) E273K*
(II) A282T*

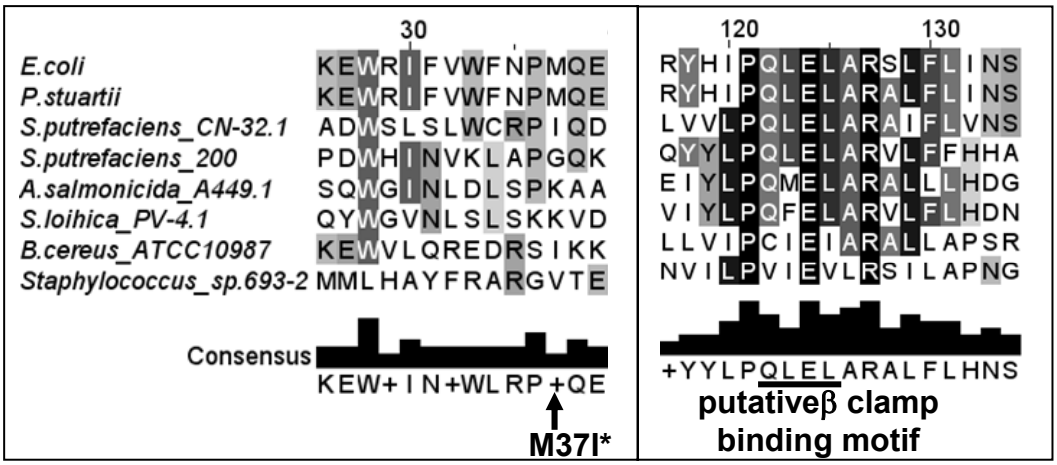

400
410
E.coli IARYSDLVVPEIDKRLIQLQ
S.putrefaciens_CN-32.1 IAQYSDLMLPEMDKRLIQLQ
S.putrefaciens_CN-32.2 LEQYDDLLMPEIEARMLSMV
S.baltica_OS155.1 IVKYSDMRVVDLDKRLIQLG
S.baltica_OS155.2 KNESKKETPNNEASEEAKVI
A.salmonicida_A449.1 IARYSDLSIPDIDKKILALS
A.salmonicida_A449.2 FEDLSPMLILESVASGHEVA
Staphylococcus_sp.693-2 MMKYEDIMINLDEVMLNHKR
H.aurantiacus_ATCC_23779 FMLM----EDIQPINLEIAL
B.cenocepacia_MC0-3 MAEFDDLLFGSEYEFLMKQL
Rhodobacter_HTCC2255 LEQCSDIVIPKIEGALLNTF
Consensus
IAQYSDLLIPEIDKRL+QL+
(II) S401F

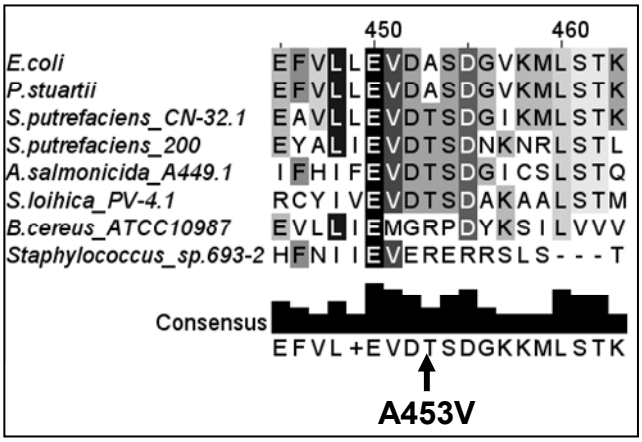
TnsC homologs contain both classes of mutation that were described in Chapter 1. Class I mutations result in untargeted transposition, but maintenance of target-site immunity. Class II mutations result in TnsC proteins are deaf to target-site immunity signals and cause random transposition. The presence of these mutations may explain why some Tn7-like transposons are not found in the *attTn7*-site, but do not explain how target-site immunity is overcome to form genomic islands, since Class II mutations do not respond to TnsD and insert into the specific *attTn7* locus.

Analysis of TnsE homologs

Mutations in TnsE that increase TnsE-mediated transposition frequency have been isolated, and most increase the DNA binding affinity of the protein (Peters and Craig, 2001). There does appear to be one increased activity mutation that does not increase DNA binding affinity, suggesting that DNA binding is not the only important aspect of TnsE (Q. Shi and J.E.Peters, unpublished observation). This increased-activity mutation (M37I) is found in the N-terminus of the TnsE protein while all mutations that increase DNA binding ability are located in the C-terminus of the protein. M37I may represent a class of mutation that alters protein-protein interactions either between TnsE and other Tns proteins or between TnsE and a host factor (see Chapter 3). Many of the mutations isolated in the laboratory can be found within the nautrally occurring *tnsE* genes, including M37I (Figure 2.8). Secondary structure predictions of the TnsE gene hint that the N-terminus and C-terminus may be separated by a mostly unstructured region between positions ~275 and ~375 (Figure 2.9). Sequence homology between TnsE proteins is very weak in this region as well (data not shown). The 275-375

Figure 2.8. Alignment of TnsE
Alignment of selected TnsE homologs that contain residues matching mutations that were isolated in screens for increased TnsABC+E transposition frequency. TnsE mutations isolated in the screen are indicated with arrows (Peters and Craig, 2001). Sites where homologs contain the exact residue recovered in mutant screens of TnsE from *E. coli* are marked with an asterisk. Note the heavy representation of mutations in the C-terminus, possibly representing a DNA binding region.


30
E.coli KEWRIFVWFNPMQE
P.stuartii KEWRIFVWFNPMQE
S.putrefaciens_CN-32.1 ADWSLSLWCRPIQD
S.putrefaciens_200 PDWHINVKLAPGQK
A.salmonicida_A449.1 SQWGINLDLSPKAA
S.loihica_PV-4.1 QYWGVNLSLSKKVD
B.cereus_ATCC10987 KEWVLQREDRSIKK
Staphylococcus_sp.693-2 MMLHAYFRARGVTE
Consensus
KEW+IN+WLRP+QE
M37I*
120
130
RYHIPQLELARSLFLINS
RYHIPQLELARALFLINS
LVVLPQLELARAIFLVNS
QYYLPQLELARVLFFHHA
EIYLPQMELARALLLHDG
VIYLPQFELARVLFLHDN
LLVIPCIEIARALLAPSR
NVILPVIEVLRSILAPNG
+YYLPQLELARALFLHNS
putativeβ clamp
binding motif


450
460
E.coli EFVLLEVDASDGVKMLSTK
P.stuartii EFVLLEVDASDGVKMLSTK
S.putrefaciens_CN-32.1 EAVLLEVDTSDGIKMLSTK
S.putrefaciens_200 EYALIEVDTSDNKNRLSTL
A.salmonicida_A449.1 IFHIFEVDTSDGICSLSTQ
S.loihica_PV-4.1 RCYIVEVDTSDAKAALSTM
B.cereus_ATCC10987 EVLLIEMGRPDYKSILVVV
Staphylococcus_sp.693-2 HFNIIEVERERRSLS---T
Consensus
EFVL+EVDTSDGKKMLSTK
A453V

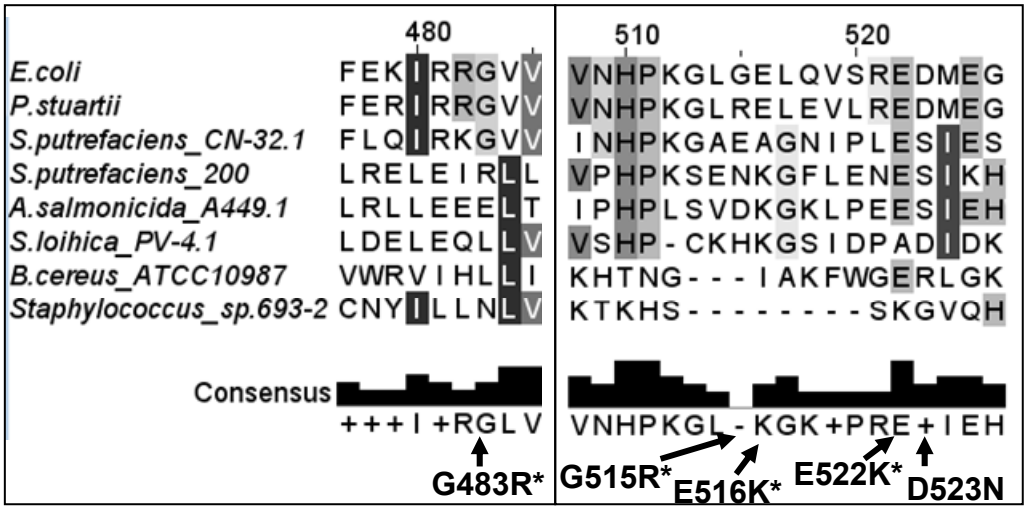
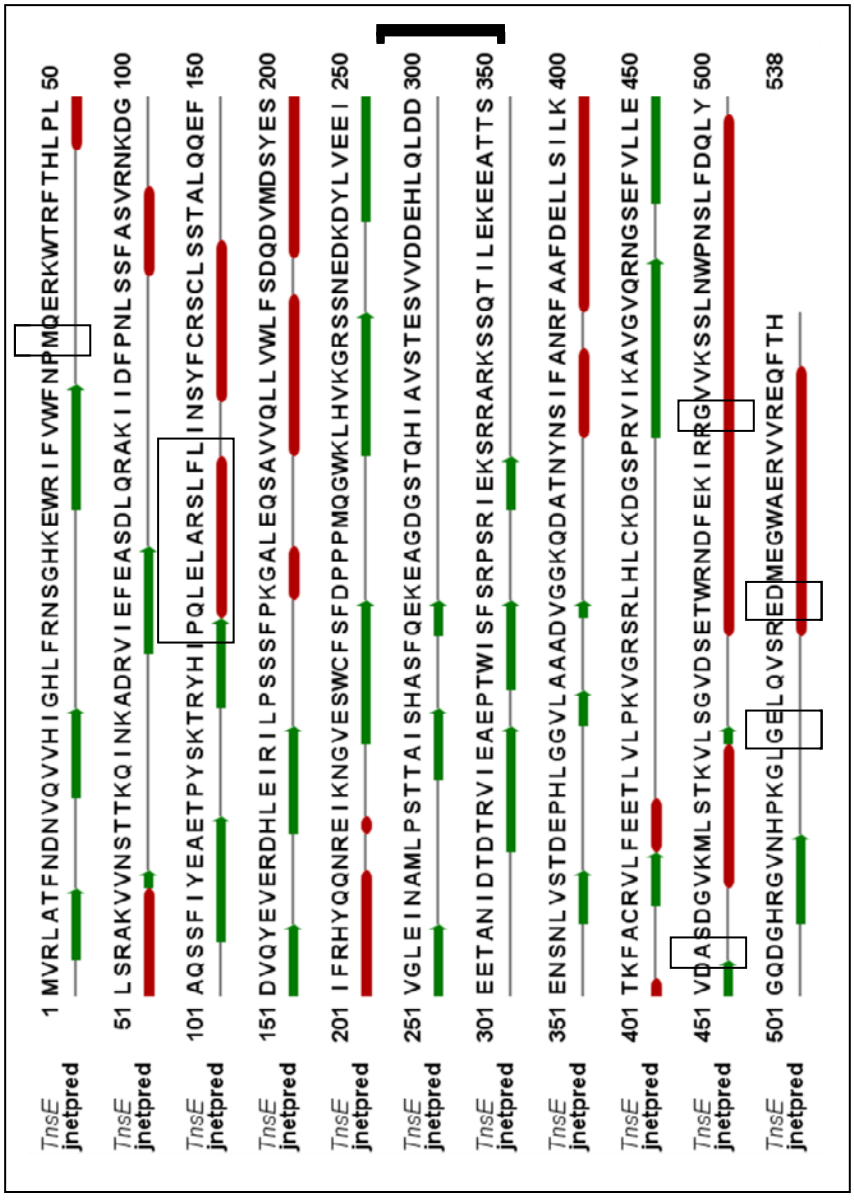

480
E.coli FEKIRRGVV
P.stuartii FERIRRGVV
S.putrefaciens_CN-32.1 FLQIRKGVV
S.putrefaciens_200 LRELEIRLL
A.salmonicida_A449.1 LRLLEEELT
S.loihica_PV-4.1 LDELEQLLV
B.cereus_ATCC10987 VWRVIHLLI
Staphylococcus_sp.693-2 CNYILLNLV
Consensus
+++I+RGLV
G483R*
510
520
VNHPKGLGELQVSREDMEG
VNHPKGLRELEVLREDMEG
INHPKGAEAGNIPLESIES
VPHPKSENKGFLENESIKH
IPHPLSVDKGKLPEESIEH
VSHP-CKHKGSIDPADIDK
KHTNG---IAKFWGERLGK
KTKHS---------SKGVQH
VNHPKGL-KGK+PRE+IEH
G515R*
E516K*
E522K*
D523N

Figure 2.9. Secondary structural prediction of TnsE
Secondary structural prediction of TnsE reveals a probable two domain structure. Arrows represent predicted beta-sheets and ovals represent predicted helical regions. The region from residues ~275 to ~375 (bracketed) is largely unstructured and poorly conserved amongst homologs of TnsE. This region is also rich in glycine and polar residues, suggesting a high degree of flexibility. Boxed residues indicate sites of mutations that are described elsewhere in this dissertation. Structural predictions were carried out with using the Jnet program (Cuff and Barton, 2000) through the Jalview softeware package (Clamp et al., 2004).


TnsE jnetpred
1 MVRLATFNDNVQVVHIGHLFRNSGHKEWRIFVWFNPMQERKWTRFTHLPL 50
51 LSRAKVVNSTTKQINKADRVIEFEASDLQRAKIIDFPNLSSFASVRNKDG 100
101 AQSSFIYEAETPYSKTRYHIPQLELARSLFLINSYFCRSCLSSTALQQEF 150
151 DVQYEVERDHLEIRILPSSSFPKGALEQSAVVQLLVWLFSDQDVMDSYES 200
201 IFRHYQQNREIKNGVESWCFSFDPPPMQGWKLHVKGRSSNEDKDYLVEEI 250
251 VGLEINAMLPSTTAISHASFQEKEAGDGSTQHIAVSTESVVDDEHLQLDD 300
301 EETANIDTDTRVIEAEPTWISFSRPSRIEKSRRARKSSQTILEKEEATTS 350
351 ENSNLVSTDEPHLGGVLAAADVGGKQDATNYNSIFANRFAAFDELLSSILK 400
401 TKFACRVLFEETLVLPKVGRSRLHLCKDGSPRVIKAVGVQRNGSEFVLLE 450
451 VDASDGVKMLSTKVLSGVDSETWRNDFEKIRRGVVKSSLNWPNSLFDQLY 500
501 GQDGHRGVNHPKGLGELQVSREDMEGWAERVVREQFTH 538

region also contains an enrichment of glycine and polar residues, suggesting that this part of the protein may constitute a solvent-exposed flexible linker that joins two separate domains, each of which may carryout different functions for the activity of TnsE. In Chapter 3, I will describe an interaction between TnsE and the processivity factor of the DNA replication machinery, the β clamp, which appears to reside in the N-terminus of.

2.4. Concluding remarks

The molecular mechanisms of Tn7 transposition contribute to the evolution of both the host genome and the transposon by maximizing the collection of genes with beneficial functions without harming the host. The ability of Tn7 to selectively direct transposition into the chromosome, mobile plasmids, and bacteriophages allows it to move in and out of stationary and mobile DNA pools while preventing the disruption of important host functions. This property of the transposon likely exposes it to a great diversity of environments in which various types of genes may be sampled and selected.

Some residues within homologs of the Tns proteins match those found in genetic screens that affect the regulation of transposition. It is difficult to say to what degree each of these transposon components conforms to the mechanisms established for Tn7 transposition without experimental evidence, but one can speculate that there is a dynamic give-and-take between mutations that diminish and restore regulation of transposition. Analysis of Tn7-like elements from different organisms should lead to a better understanding of the evolutionary constraints that maintain regulation of these transposons, and possibly recombination in general.

While transposons are rightly considered “selfish” genetic elements that

parasitize the genome of organisms to ensure their own propagation, Tn7 may tell a slightly different story. The highly evolved targeting system of Tn7 virtually ensures that deleterious insertions do not occur while allowing the host to sample genes collected through a variety of mobile DNAs. The existence of genomic islands within *attTn,7* where little or no Tn7 transposition functions remain, underscores the ability of the host to maintain genetic information based exclusively on its value to the host.

2.5 Experimental procedures

Phylogenetic analysis of Tn7-like elements

The predicted amino acid sequences of TnsABCD from each element was concatenated and aligned using the Jalview software package and the ClustalW algorithm (see Table S1 for accession numbers) (Chenna et al., 2003; Clamp et al., 2004). Protein distances and Neighbor Joining trees were constructed using the Phylip program (Felsenstein, 1993). The tree was rooted on *Acidithiobacillus ferroxidans*. Bootstrap analysis was also carried out using 100 iterations, and values greater than 80% are given at appropriate nodes. The phylogenetic tree was drawn and manipulated using the MEGA4 software package (Tamura et al., 2007).

Identification of Tn7 ends and determination of the location of attTn7 loci

Sequences betwee the *glmS* and *tnsA* genes were analyzed for tandem repeats by visual inspection using Artemis (Rutherford et al., 2000) and computationally using the etandem program of the EMBOSS software package (Rice et al., 2000). Artemis Comparison Tool (release 4) was used to compare Tn7 containing species to non-Tn7 containing species to aid in the

location of left ends and in defining empty *attTn7* loci (Abbott et al., 2005). TnsB binding sites that were determined by analysis of the right end (between *glmS* and *tnsA*) were used to search for the more degenerate and widely spaced TnsB binding sites of the left end.

Analysis of genes transported by Tn7-like elements

The function of genes found between the ends of the Tn7-like elements were surmised using the BLAST algorithm to search gene databases for related genes with known functions (Altschul et al., 1990; Benson et al., 2005; Peterson et al., 2001). Genes were then grouped by functional categories of. Functional categories of genes are loosely modeled after CMR-TIGR functional categories (Peterson et al., 2001).

Alignment of predicted protein sequences

The BLAST and PSI-BLAST algorithms were used to search gene databases for TnsABCDE homologs using sequences from Tn7 from *E. coli* as the query (Altschul et al., 1990; Altschul et al., 1997; Benson et al., 2005; Peterson et al., 2001). The predicted amino acid sequences of TnsA, TnsB, TnsC, and TnsE from each element were aligned using the Jalview software package and the ClustalW algorithm (see Table 2.1 for accession numbers) (Chenna et al., 2003; Clamp et al., 2004).

Secondary structure prediction of TnsE

Structural predictions were carried out with using the Jnet program (Cuff and Barton, 2000) through the Jalview software package (Clamp et al., 2004).

2.6 Acknowledgements

I am grateful to Zaoping Li and Qiaojuan Shi for comments on the manuscript and to Steve Zinder for advice on issues of phylogeny. This work was funded by the National Institutes of Health grant GM069508.

REFERENCES

Abbott, J. C., Aanensen, D. M., Rutherford, K., Butcher, S., and Spratt, B. G. (2005). WebACT--an online companion for the Artemis Comparison Tool. Bioinformatics *21*, 3665-3666.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Arciszewska, L. K., and Craig, N. L. (1991). Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. Nucl Acids Res *19*, 5021-5029.

Barlow, R. S., and Gobius, K. S. (2006). Diverse class 2 integrons in bacteria from beef cattle sources. J Antimicrob Chemother *58*, 1133-1138.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2005). GenBank. Nucleic Acids Res *33*, D34-38.

Biskri, L., and Mazel, D. (2003). Erythromycin esterase gene ere(A) is located in a functional gene cassette in an unusual class 2 integron. Antimicrob Agents Chemother *47*, 3326-3331.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res *31*, 3497-3500.

Choi, K. H., DeShazer, D., and Schweizer, H. P. (2006). mini-Tn7 insertion in bacteria with multiple glmS-linked attTn7 sites: example *Burkholderia mallei* ATCC 23344. Nat Protoc *1*, 162-169.

Choi, K. H., Gaynor, J. B., White, K. G., Lopez, C., Bosio, C. M., Karkhoff-Schweizer, R. R., and Schweizer, H. P. (2005). A Tn7-based broad-range bacterial cloning and expression system. Nat Methods *2*, 443-448.

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java alignment editor. Bioinformatics *20*, 426-427.

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Cuff, J. A., and Barton, G. J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins *40*, 502-511.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) (Seattle, Distributed by the author).

Godde, J. S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol *62*, 718-729.

Hansson, K., Sundstrom, L., Pelletier, A., and Roy, P. H. (2002). IntI2 integron integrase in Tn7. Journal of Bacteriology *184*, 1712-1721.

Iacono, M., Villa, L., Fortini, D., Bordoni, R., Imperi, F., Bonnal, R. J., Sicheritz-Ponten, T., De Bellis, G., Visca, P., Cassone, A., and Carattoli, A. (2008). Whole genome pyrosequencing of an epidemic multidrug resistant *Acinetobacter baumannii* of the European clone II. Antimicrob Agents Chemother.

Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res *29*, 3742-3756.

Kobayashi, I. (2004). Genetic Addiction: a Principle of Gene Symbiosis in a Genome. Plasmid Biology, 105–144.

Kuduvalli, P. N., Mitra, R., and Craig, N. L. (2005). Site-specific Tn7 transposition into the human genome. Nucleic Acids Res *33*, 857-863.

Lawrence, J. G., and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol *44*, 383-397.

Lu, F., and Craig, N. L. (2000). Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. EMBO J *19*, 3446-3457.

McKenzie, G. J., and Craig, N. L. (2006). Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. BMC Microbiol *6*, 39.

Nelson, K., Clayton, R., Gill, S., Gwinn, M., Dodson, R., Haft, D., Hickey, E., Peterson, J., Nelson, W., and Ketchum, K. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature *399*, 323-329.

Orman, B. E., Pineiro, S. A., Arduino, S., Galas, M., Melano, R., Caffer, M. I., Sordelli, D. O., and Centron, D. (2002). Evolution of multiresistance in nontyphoid salmonella serovars from 1984 to 1998 in Argentina. Antimicrob Agents Chemother *46*, 3963-3970.

Parks, A. R., and Peters, J. E. (submitted). Tn7 elements: Engendering diversity from chromosomes to episomes. Plasmid.

Parks, A. R., and Peters, J. E. (2007). Transposon Tn7 is widespread in diverse bacteria and forms genomic islands. J Bacteriol *189*, 2170-2173.

Peters, J. E., and Craig, N. L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol Cell *6*, 573-582.

Peters, J. E., and Craig, N. L. (2001). Tn7 recognizes target structures associated with DNA replication using the DNA binding protein TnsE. Genes & Dev *15*, 737-747.

Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., and White, O. (2001). The Comprehensive Microbial Resource. Nucleic Acids Res *29*, 123-125.

Quiroga, C., Roy, P. H., and Centron, D. (2008). The S.ma.I2 class C group II intron inserts at integron attC sites. Microbiology *154*, 1341-1353.

Ramirez, M. S., Quiroga, C., and Centron, D. (2005a). Novel rearrangement of a class 2 integron in two non-epidemiologically related isolates of *Acinetobacter baumannii*. Antimicrob Agents Chemother *49*, 5179-5181.

Ramirez, M. S., Vargas, L. J., Cagnoni, V., Tokumoto, M., and Centron, D. (2005b). Class 2 integron with a novel cassette array in a *Burkholderia cenocepacia* isolate. Antimicrob Agents Chemother *49*, 4418-4420.

Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E., Tettelin, H., Fouts, D. E., Eisen, J. A., Gill, S. R.*, et al.* (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature *423*, 81-86.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet *16*, 276-277.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics *16*, 944-945.

Shi, Q., Parks, A.R., Potter, B.D., Safir, I.J., Luo, Y., Forster, B.M., Peters, J.E. (2008). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in E. coli. Genetics *In Press*.

Skelding, Z., Queen-Baker, J., and Craig, N. L. (2003). Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. Embo J *22*, 5904-5917.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol *6*, 181-186.

Stellwagen, A., and Craig, N. L. (1997). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein which activates the bacterial transposon Tn7. Genetics *145*, 573-585.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol *24*, 1596-1599.

Vallenet, D., Nordmann, P., Barbe, V., Poirel, L., Mangenot, S., Bataille, E., Dossat, C., Gas, S., Kreimeyer, A., Lenoble, P.*, et al.* (2008). Comparative analysis of *Acinetobacters*: three genomes for three lifestyles. PLoS ONE *3*, e1805.

CHAPTER 3. TRANSPOSITION INTO REPLICATING DNA OCCURS THROUGH INTERACTION WITH THE PROCESSIVITY FACTOR[3]

3.1 Summary

The bacterial transposon Tn7 directs transposition into actively replicating DNA by a mechanism involving the transposon-encoded protein TnsE. Here we show that TnsE physically and functionally interacts with the processivity factor of the DNA replication machinery. This interaction allows Tn7 to identify insertion sites and to orient in one direction with active DNA replication. This work defines a new variant of the bacterial processivity factor interaction motif found within the TnsE proteins that coordinates transposition with replication. The TnsE interaction with an essential and conserved component of the replication machinery reveals a new mechanism by which Tn7, and possibly other elements, select target sites associated with DNA replication. These results provide insight into various processes in eukaryotic and prokaryotic organisms involving processivity factors.

3.2 Introduction

Target-site selection is a prerequisite for activation of transposition; transposon excision and insertion does not occur until an appropriate target has been identified. TnsD recognizes a specific site, called its attachment site

[3] *This chapter has been submitted for publication and was a collaborative effort with contributions from Qiaojuan Shi and Joseph Peters. Figure 3.6 is the result of work carried out by Q.S. The remainder of this chapter was carried out by A.R.P.

or *attTn7*, by binding to a highly conserved DNA sequence within the 3' end of the *glmS* gene. The TnsE protein recognizes an incompletely defined feature associated with discontinuous DNA replication (Peters and Craig, 2001a) that is over-represented or especially accessible in mobile plasmids, called conjugal plasmids, as they enter a new host cell (Wolkow et al., 1996).

While TnsE-mediated transposition preferentially occurs into mobile plasmids undergoing conjugal DNA replication, at a lower frequency, the TnsABC+E machinery also recognizes sites within the bacterial chromosome with a preference for the region where DNA replication terminates and regions proximal to DNA double-strand breaks (Peters and Craig, 2000). The orientation of the transposon ends resulting from TnsE-mediated insertions indicates that discontinuously replicated DNA is in some way recognized by TnsE (Peters and Craig, 2001a; Wolkow et al., 1996). As mobile plasmids enter a new host cell, they replicate in a single direction by a discontinuous process, like lagging-strand DNA synthesis (Wilkins and Lanka, 1993). In mobile plasmids and in the chromosome, transposition events occur in a single orientation correlating with the direction of replication fork progression (Peters and Craig, 2001a; Peters and Craig, 2001b; Wolkow et al., 1996). It has been shown that TnsE is a DNA binding protein that preferentially binds to DNA structures that present a free 3' recessed end (Peters and Craig, 2001b). Given that TnsD relies in part on additional host factors in activating transposition (Sharpe and Craig, 1998), it is conceivable that host factor associated with discontinuous DNA replication allows the selection of targets in TnsE-mediated transposition.

An intriguing host factor candidate that could allow the TnsABC+E transposition machinery to target lagging-strand DNA synthesis is the DNA

replication processivity factor. Interestingly, the inactive *pogo* element, found in *Drosophila*, encodes a transposase that has been shown to interact with the processivity factor (Warbrick et al., 1998). However, because the element is no longer active, no functional link between this interaction and transposition has yet been determined (Warbrick, 2000; Warbrick et al., 1998). It therefore remains unclear if this interaction was artifactual or, if the interaction was important in the original element, whether it regulated transposition with DNA replication or repair, or was involved in target-site selection. We reasoned that TnsE might use an interaction with the processivity factor to direct transposition into certain forms of DNA replication.

The processivity factor, β clamp in Bacteria and PCNA in Eukaryotes and Archaea, is enriched on discontinuously replicating DNA. Processivity factors are clamp proteins that encircle DNA and serve as a mobile platform, linking proteins to DNA (reviewed in Johnson and O'Donnell, 2005, and references therein). β and PCNA have been shown to interact with a number of proteins involved in DNA repair, Okazaki fragment maturation, and regulation of the cell cycle (Johnson and O'Donnell, 2005; Warbrick, 2000). The processivity factor binding motif found in proteins that interact with β and PCNA is commonly found in either the extreme C- or N-terminus of the protein and fits into a hydrophobic cleft of the clamp proteins (Dalrymple et al., 2001; Johnson and O'Donnell, 2005; Warbrick, 2000). Competition for the common binding site appears to play a role in coordination of proteins involved in DNA metabolism (Lopez de Saro et al., 2004). Studies have shown that proteins can exclude the binding some interacting partners, but not others, suggesting that there may be a hierarchy or a sequence in which proteins are allowed to bind to the processivity factor (Sutton, 2004; Sutton and Duzen, 2006; Sutton

et al., 2005). In *E. coli*, β is not produced in high enough quantities to complete an entire round of chromosomal replication without recycling clamps off of completely replicated DNA behind the replication fork back into the replisome for use in ongoing replication. The necessity to recycle clamps may provide a further level of regulation involving processivity factor usage.

The processivity factors are loaded onto DNA by a multi-subunit protein complex called the gamma complex (in bacteria) or the RFC complex (in eukaryotes) in an ATP dependent manner (Johnson and O'Donnell, 2005; Warbrick, 2000). In bacteria, the complex that is necessary and sufficient to load the β clamp onto DNA consists of γ (or τ), δ, and δ' subunits. It is thought that the δ subunit alone is sufficient to remove the β clamp from DNA, allowing it to be used again for further replication. δ is found in molar excess of the remaining subunits, and has been shown in vitro to remove β from DNA in the absence of ATP (Leu et al., 2000). ATP is required for γ and δ' to regulate the ability of δ to interact with the hydrophobic interface between the β homo-dimer, which is the principle means of the clamp open property of δ (Johnson and O'Donnell, 2005). The gamma complex requires a free 3' DNA end, such as a nick or gapped structure, to load clamps onto double-stranded DNA.

The central role of the β clamp in DNA replication, particularly in regulation of proteins on the lagging-strand, make interaction with this protein an attractive component of a model for explaining how TnsE identifies lagging-strand DNA synthesis. In this Chapter, we reveal and characterize an interaction between TnsE and the processivity clamp associated with DNA replication. We show that this interaction is essential for TnsE activity in vivo and find that TnsE has evolved to interact with clamps not actively involved in an essential step in chromosomal DNA replication. These findings likely

reveal a general strategy used by multiple transposons for directing transposition to DNA replication intermediates.

3.4 Results

TnsE interacts with the processivity factor

In an effort to understand how TnsE identifies a target DNA we looked for conserved motifs within the amino acid sequence of TnsE. We found a sequence that shows a modest resemblance to the consensus processivity factor binding motif found in bacterial host proteins (QL(S/D)LF) (Dalrymple et al., 2001)(Figure 3.1). Analysis of amino acid alignments of all known and predicted *tnsE* gene products using the ClustalW algorithm (Thompson et al., 1994) revealed a highly conserved sequence PQLELARALFL (Figure 3.1, Figure 2.8). We hypothesized that TnsE recognizes lagging-strand DNA synthesis through an interaction with the processivity factor. We assessed the TnsE-β interaction using TnsE and β protein derivatives fused to the yeast transcription activation and DNA binding domains, respectively (Fields and Song, 1989). The presence and extent of the interaction in the two-hybrid assay was monitored by determining the β-galactosidase (β-gal) activity in a reporter strain containing a *lexA*::*lacZ* fusion (Liachko and Tye, 2005). We also included a positive control for the β interaction, the δ subunit of the clamp loader. The yeast two-hybrid assay indicated that TnsE and β interact (Figure 3.2.). To test if the putative β-clamp binding motif in TnsE is important for binding to β, we constructed a set of *tnsE* mutants replacing specific amino acids within this region with alanines (*tnsE*$^{\beta MA}$). Amino acids were chosen for mutation based on their presumed relationship to the consensus β clamp binding motif and their conservation among TnsE proteins (Figure 3.1). Each

Figure 3.1. Alignment of TnsE homologs reveals a putative β clamp binding motif. A representation of the ~538 amino acid TnsE protein found in *E. coli* is shown with the amino (N) and carboxy (C) termini indicated. An alignment of TnsE homologs encompassing the putative β clamp binding motif is presented with the putative motif boxed. The consensus sequence of the region between residues 121-131 containing the putative β clamp binding motif are shown in bold. Underlined residues correspond to positions where alanine substitutions were made in the $TnsE^{\beta MA}$ mutants. The consensus β clamp binding motif found in bacterial host proteins is commonly reported as QL(S/D)LF (Dalrymple et al., 2001). Alignment and consensus sequence determination of predicted TnsE proteins that were described previously (see chapter 2)(Parks and Peters, 2007), was performed using the ClustalW application (Thompson et al., 1994) in the Jalview software (Clamp et al., 2004).

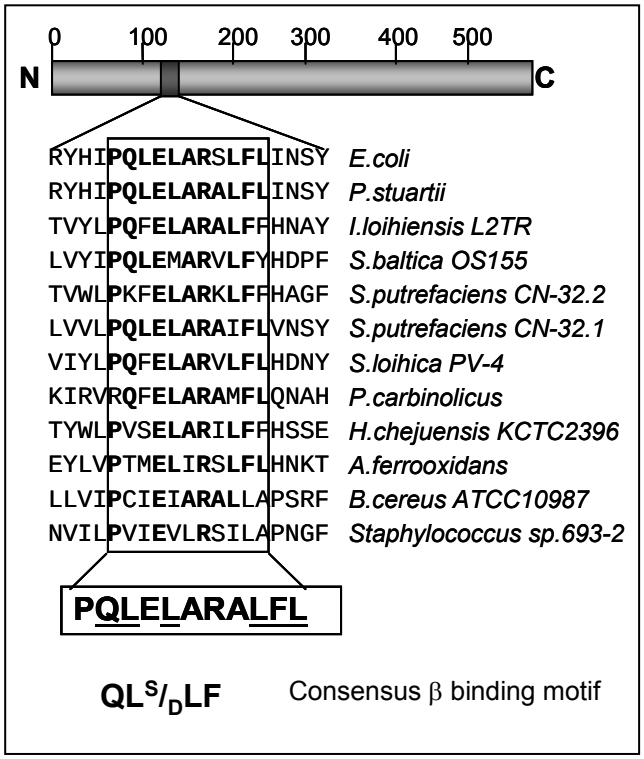
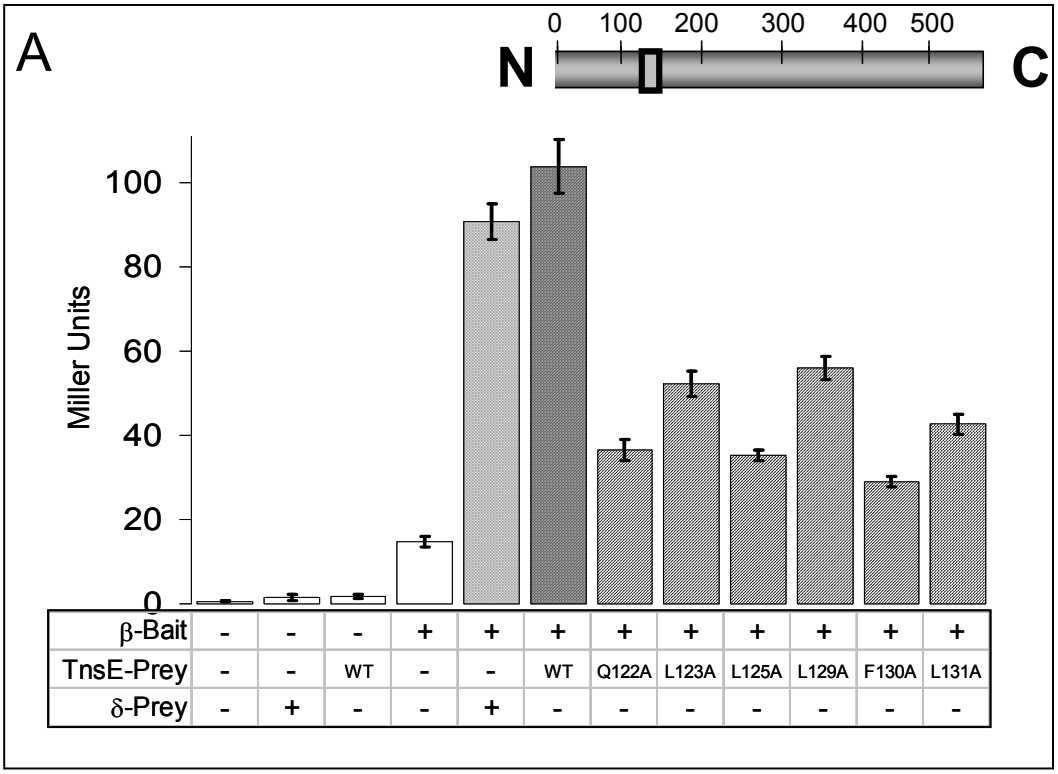
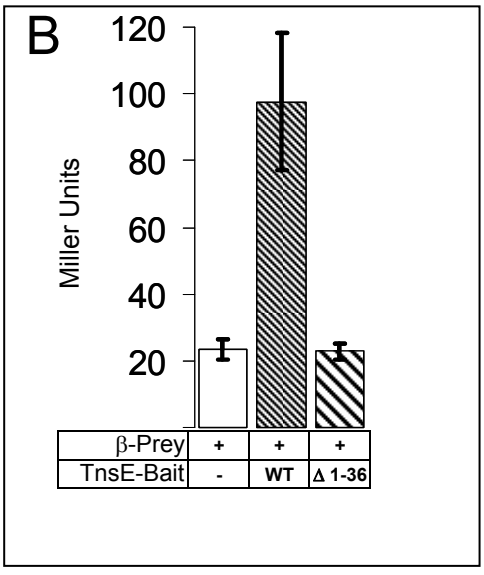

0
100
200
300
400
500
N
C
RYHIPQLELARSLFLINSY E.coli
RYHIPQLELARALFLINSY P.stuartii
TVYLPQFELARALFFHNAY I.loihiensis L2TR
LVYIPQLEMARVLFYHDPF S.baltica OS155
TVWLPKFELARKLFFHAGF S.putrefaciens CN-32.2
LVVLPQLELARAIFLVNSY S.putrefaciens CN-32.1
VIYLPQFELARVLFLHDNY S.loihica PV-4
KIRVRQFELARAMFLQNAH P.carbinolicus
TYWLPVSELARILFFHSSE H.chejuensis KCTC2396
EYLVPTMELIRSLFLHNKT A.ferrooxidans
LLVIPCIEIARALLAPSRF B.cereus ATCC10987
NVILPVIEVLRSILAPNGF Staphylococcus sp.693-2
PQLELARALFL
QLS/DLF
Consensus β binding motif

Figure 3.2. Yeast two-hybrid analysis reveals an interaction between TnsE and β. **A**. Quantification of the yeast two-hybrid assay reveals a defect in the β interaction with TnsE proteins with alanine substitutions in the putative β interaction motif. For the assay the bacterial proteins were fused to either the yeast DNA binding domain (Bait) or the yeast transcription activation domain (Prey)(Liachko and Tye, 2005). The β-fusion alone displays a slight auto-activation effect, while TnsE and the positive control, the δ subunit of the clamp loader, display interaction signals significantly above background. The alanine substitutions in the putative β clamp binding motif consistently reduce TnsE-β interaction levels. Interaction was measured by Miller assay and is reported in Miller Units(Miller, 1992). **B**. Comparison of β-galactosidase activity produced in the presence of wild-type fusion proteins with Δ1-36 reveals an apparent loss of interaction with β in the absence of the N-terminus. Error bars indicate standard error of the mean (n=4).


A
0
100
200
300
400
500
N
C
Miller Units
100
80
60
40
20
0
β-Bait - - - + + + + + + + + +
TnsE-Prey - - WT - - WT Q122A L123A L125A L129A F130A L131A
δ-Prey - + - - + - - - - - - -

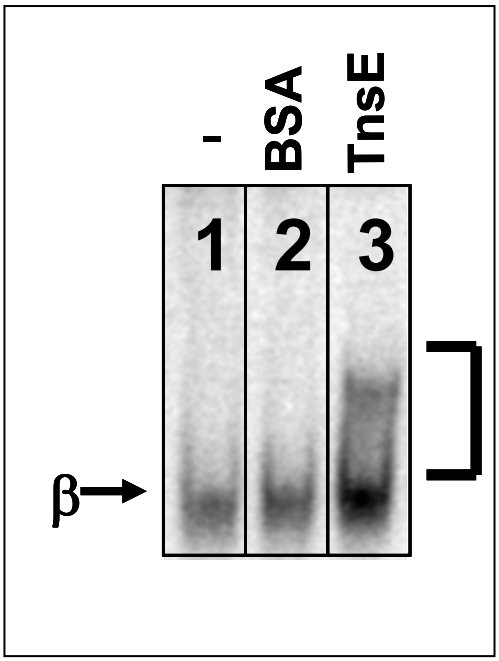

B
120
100
80
60
40
20
Miller Units
β-Prey + + +
TnsE-Bait - WT Δ 1-36

alanine substitution resulted in significantly decreased β-gal activity ($p<0.0005$, two-tailed unequal variance t-test) in the reporter strain, supporting the idea that the conserved region is important for interaction with β and possibly comprises a region that interacts with the hydrophobic pocket of the β clamp (Figure 3.2.A.). Western blots do not show changes in stability or expression resulting from the $TnsE^{\beta MA}$ mutations that would account for these observations (data not shown).

The *tnsE* gene begins with a rare valine start codon (Peters and Craig, 2001a). Experiments have shown that deletion of residues 1-36 residues that precede the first ATG codon yields an inactive protein, and that this protein does not act in a dominant negative fashion (Shi and Peters, unplublished observation). We tested to see if these first 36 amino acids are involved in interaction with β using the yeast two-hybrid assay. We found that interaction signal with β is returned to background when residues 1-36 are truncated (Figure 3.2.B). This result is consistent with the idea that interaction with β correlates with TnsE activity, and that this interaction is associated with the N-terminus of the TnsE protein.

To confirm the interaction between TnsE and β in vitro, we purified a modified β protein that could be labeled with ^{32}P phosphate (^{32}P-β) (Kelman et al., 1995a; Kelman et al., 1995b). We tested the TnsE-β interaction using a protein mobility shift assay (Lopez de Saro et al., 2006; Lopez de Saro and O'Donnell, 2001). A shift in the electrophoretic mobility of ^{32}P-β upon addition of TnsE demonstrated that TnsE and β do form a complex that is robust enough to be observed under the assay conditions (Figure 3.3). To determine relative binding abilities of $TnsE^{wt}$ and a subset of the $TnsE^{\beta MA}$ mutants we used a far western blot technique (Einarson et al., 2007). Serial dilutions of

Figure 3.3. Protein mobility shift assays confirm the TnsE-β interaction in vitro. ^{32}P-labeled β monomer (50 nM) (^{32}P-β alone in lane 1) is unaffected by the addition of 2.0 μM BSA (lane 2) in a 4% native polyacrylamide gel, but produces a shifted product (black bracket) with the addition of 2.0 μM TnsE^{wt} (lane 3). Protein gel shift assays were conducted as described (Lopez de Saro et al., 2006; Lopez de Saro and O'Donnell, 2001) with proteins purified as previously described(Kelman et al., 1995a; Kelman et al., 1995b; Peters and Craig, 2001a).The membrane was then probed with ^{32}P-β. Consistent with the yeast two-hybrid data, we found that TnsE and β.


-
BSA
TnsE
1
2
3
β

$TnsE^{wt}$ and the $TnsE^{\beta MA}$ mutants were spotted on a membrane in triplicate. interact and that the TnsE proteins containing alanine substitutions in the putative β-binding motif were compromised for β clamp interaction (Figure 3.4).

The dissociation constant (Kd) of a complex is a measure of how tightly two proteins interact with one another. We sought to determine the dissocitation constant of the TnsE-β complex using an equilibrium gel filtration technique (Beeckmans, 1999). In this technique, a gel matrix is equilibrated with a buffer including one of the proteins (^{32}P-β). The test protein (TnsE) is run on the column in the same buffer containing ^{32}P-β. The TnsE-β complex elutes ahead of the individual proteins, due to its larger size and exclusion from the matrix, and is observed as a peak in autoradiography. The fractions that usually contain β alone are depleted and appear as a trough. Since the concentration of ^{32}P-β is constant throughout the experiment , the system remains at equilibrium and an accurate measure of Kd can be determined by evaluating the area under the peak (Figure 3.5). We carried out this experiment with two different concentrations of ^{32}P-β (50 nM and 1 nM) and two amounts of TnsE (790 pmole and 200 pmole). Surprisingly we observed four distinct peaks, possibly indicating oligomeric states of the TnsE-β complex. When combined, the peaks gave a Kd value of 400 nM. Since it is unclear what the stoichiometry of binding is, this value is a crude estimate. This Kd value would be consistent with the relatively weak binding that we expected to see, and was intermediate between the interaction observed

Figure 3.4 Far western blot with $TnsE^{WT}$ and $TnsE^{\beta MA}$ mutants
Wild type TnsE and a subset of three $TnsE^{\beta MA}$ mutants are shown with ^{32}P signal plotted vs. amount of protein spotted on the membrane. Each point represents the average from three experiments. A general trend of decreased binding ability from each of the mutants can be seen at all concentrations. $TnsE^{WT}$ is represented with solid diamonds (♦), $TnsE^{Q122A}$ with open circles (o), $TnsE^{L123A}$ with open triangles (Δ), and $TnsE^{L125A}$ with open squares (□). Far western blot assays were carried out as described (Einarson et al., 2007) using Immobilon-P membranes (Millipore).

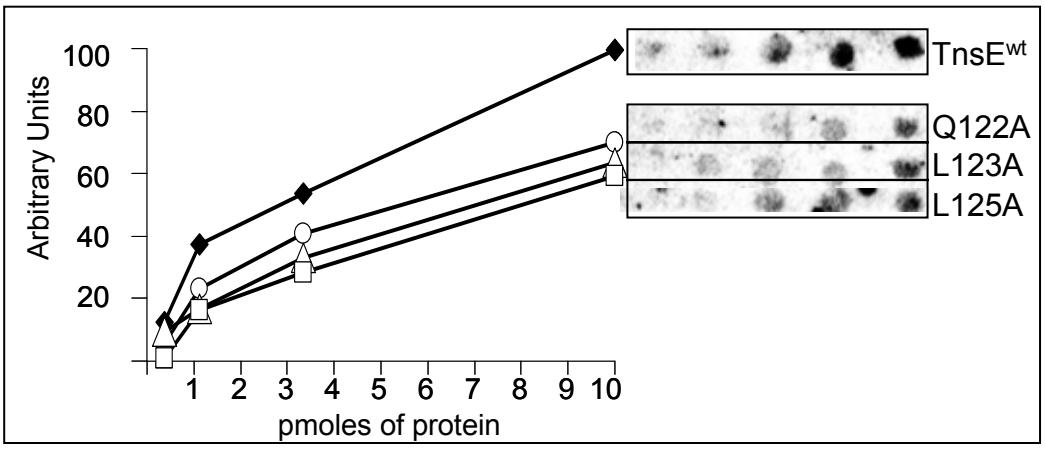
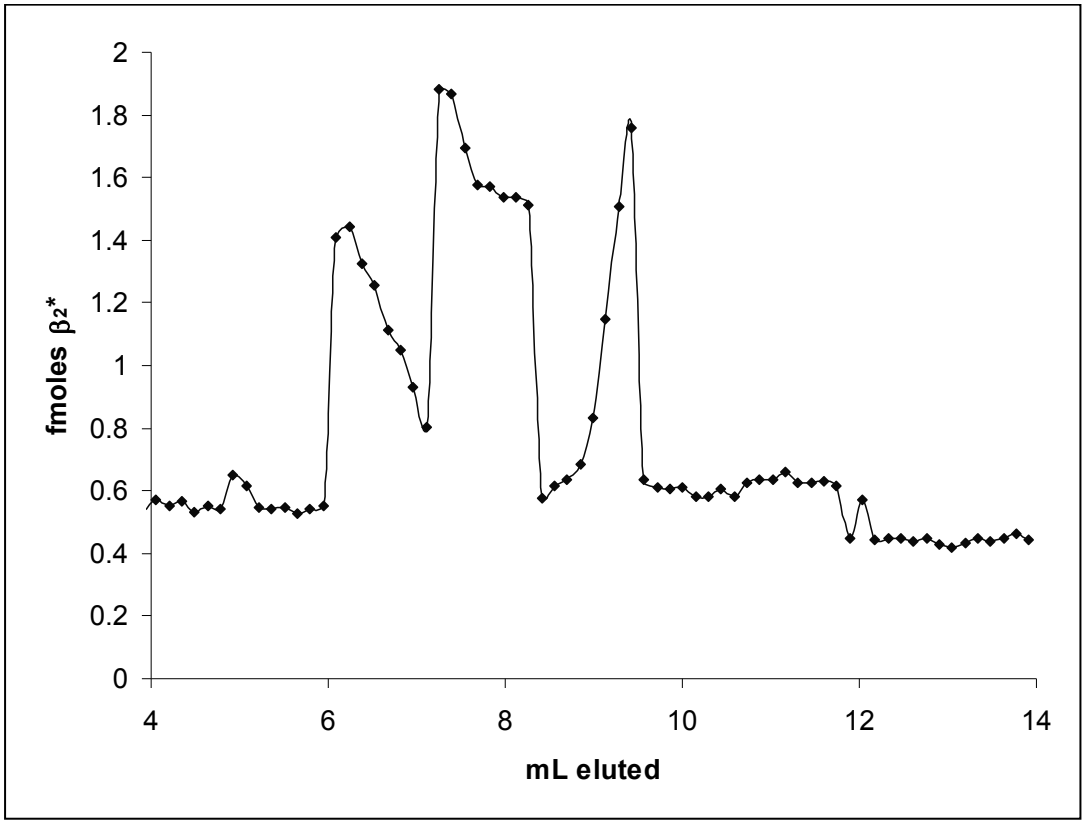

100
80
60
40
20
Arbitrary Units
1 2 3 4 5 6 7 8 9 10
pmoles of protein
TnsE^wt
Q122A
L123A
L125A

Figure 3.5 Equilibrium gel filtration analysis of TnsE-β interaction
An equilibrium gel filtration technique provides further evidence for a β-TnsE complex. We see multiple peaks, possibly indicating different oligomeric states of TnsE-β complexes. A 10 ml column was packed with 4% agarose beads and equilibrated with buffer containing 1nM ^{32}P-β. 790 pmoles of TnsE was loaded onto the column in the same buffer containing 1nM ^{32}P-β. 145ul fractions were eluted from the column, and 50ul of each fraction was analyzed by liquid scintillation counter. This plot shows the number of fmoles of ^{32}P-labeled β at a given total elution volume. Peaks indicate TnsE-β complexes, while the trough is evidence of depletion of ^{32}P-β due to complex formation.

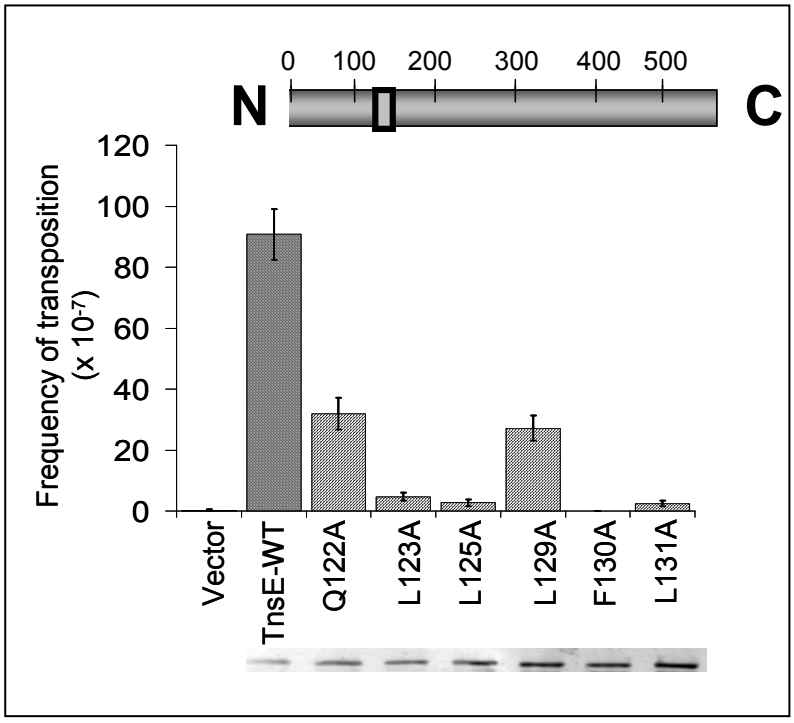

2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
fmoles β2*
4
6
8
10
12
14
mL eluted

between the delta (30 nM) and gamma (900 nM) subunits of the clamp loader complex. Additional optimization and controls are required to more conclusively determine the Kd of the TnsE-β complex.

Interaction with the processivity factor is essential for transposition in vivo

If binding to the β clamp is required to activate transposition, we expected to see a decrease in transposition with the *tnsE*$^{\beta MA}$ mutants, due to their attenuated β clamp binding ability. Interaction with the clamp via the interaction motif is required for activity in vivo for DNA polymerases (Becherel et al., 2002; Beuning et al., 2006; Lenne-Samuel et al., 2002), although these protein-protein interactions include more extensive surface interactions beyond the conserved motif (Bunting et al., 2003). Using an in vivo transposition assay (McKown et al., 1988), we found that the alanine mutations abolished or significantly reduced ($p<0.005$, two-tailed unequal variance t-test) the frequency of TnsABC+E transposition (Figure 3.6). Western blots confirmed that the stability and expression of the mutants did not account for this result (data not shown). The decreases we observed in transposition frequency are consistent with decreased β clamp binding ability, supporting the view that activation of transposition via the TnsE pathway is dependent on binding to the β clamp. Similarly, the N-terminally truncated TnsE mutant protein does not mediate transposition and shows decreased interaction in the Yeast two-hybrid assay.

Figure 3.6. Transposition frequency with TnsE^{WT} and TnsE$^{\beta MA}$ mutants
Transposition frequency is reduced significantly below wild type with each of the six TnsE$^{\beta MA}$ mutations. Transposition was monitored in cells expressing TnsABC and wild type TnsE or a mutant TnsE containing an alanine substitution at one of six positions in the putative β clamp interacting motif (see text for details). Transposition assays were conducted in *recA*$^-$ cells containing *tns* genes on plasmids using a lambda delivery vector (McKown et al., 1988). Error bars indicate the standard error of the mean (n=3). This analysis was carried out by Qiaojuan Shi. Western blots against TnsE show similar levels of TnsE are produced from each TnsE containing vector.


0
100
200
300
400
500
N
C
120
100
80
60
40
20
0
Frequency of transposition
(x 10-7)
Vector
TnsE-WT
Q122A
L123A
L125A
L129A
F130A
L131A

We hypothesized that if TnsE relies on the β clamp for target recognition we might increase the probability of TnsE recognizing a target by increasing the concentration of β in the cell. Increasing the concentration of β in the cell may interfere with the cycling of β clamps off DNA resulting in an increased retention time of β clamps on completed Okazaki fragments. The δ subunit of the processivity factor clamp loader is present in molar excess of the entire complex (γ-complex) within cells, and has been shown to remove clamps from DNA in vitro in the absence of other γ-complex components (Leu et al., 2000). By increasing the concentration of free β clamp in the cell, δ may be sequestered and unable to remove clamps from DNA efficiently. We performed a transposition assay with moderate over-expression of β and found that the transposition frequency doubled, compared to the empty vector control (Figure 3.7). Taken together with the data presented above, we conclude that the responsiveness of TnsABC+E transposition to the concentration of β is consistent with a dependence on interaction with β for target-site selection and subsequent activation of transposition.

We also found that mutation of *dnaN* affects TnsABC+E transposition. The protein produced by the *dnaN159* allele ($dnaN^{G66E,G174A}$) displays impaired interaction with a subset of its binding partners including the α catalytic subunit of PolIII, the δ subunit of the clamp loader, and Pol IV (Maul et al., 2007; Sutton, 2004). We hypothesized that if competition for binding to the β-clamp governs transposition, we should see an increase in transposition in strains where proteins that normally bind to the clamp are impaired for binding. We found that in *dnaN159* cells TnsABC+E transposition was double that observed in the wild-type (Figure 3.8), indicating that TnsE's ability to compete for access to the clamp may be enhanced in the *dnaN159* background. The

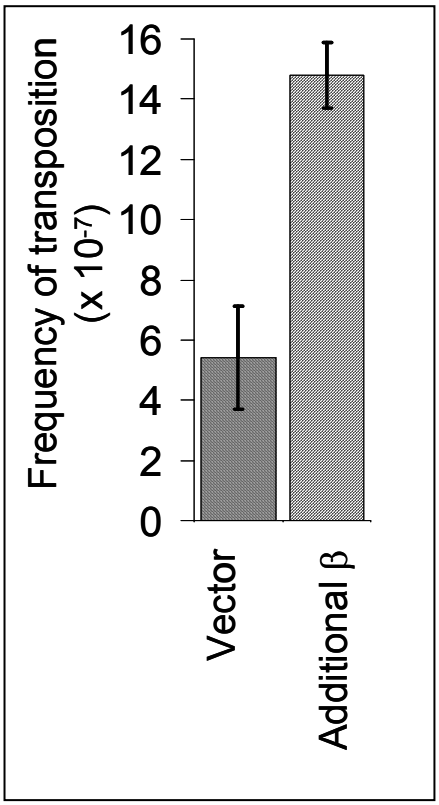


Figure 3.7. Over-expression of β results in a significant increase in TnsE-mediated transposition. Transposition was monitored in cells expressing TnsABC+E with a plasmid expressing β or an empty vector. Error bars indicate standard error of the mean (n=3).

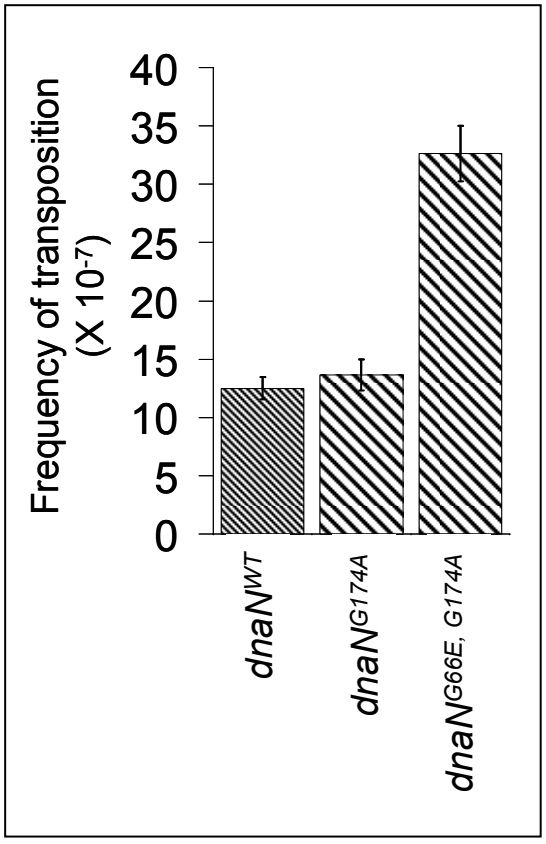


Figure 3.8. TnsE-mediated transposition frequency in mutant *dnaN* backgrounds. Presence of the *dnaN159* (G66E, G174A) allele significantly increases the frequency of TnsE-mediated transposition frequency above levels observed in wild-type cells . The *dnaN781* (G174A) allele does not stimulate transposition. This effect may be due to an increased ability for TnsE to compete with other proteins for access to the processivity clamp. Error bars indicate standard error of the mean (n=3).

dnaN781 allele contains the single mutation *dnaN*^{G174A} and does not stimulate TnsE-mediated transposition (Figure 3.8). This *dnaN781* allele displays improved interaction with PolII, PolIII, and PolIV, and is not tempertature sensitive like *dnaN159* (Maul et al., 2007). The improved interaction characteristics of host proteins likely improve coordination of proteins on the clamp and prevents prolonged exposure of free β clamps.

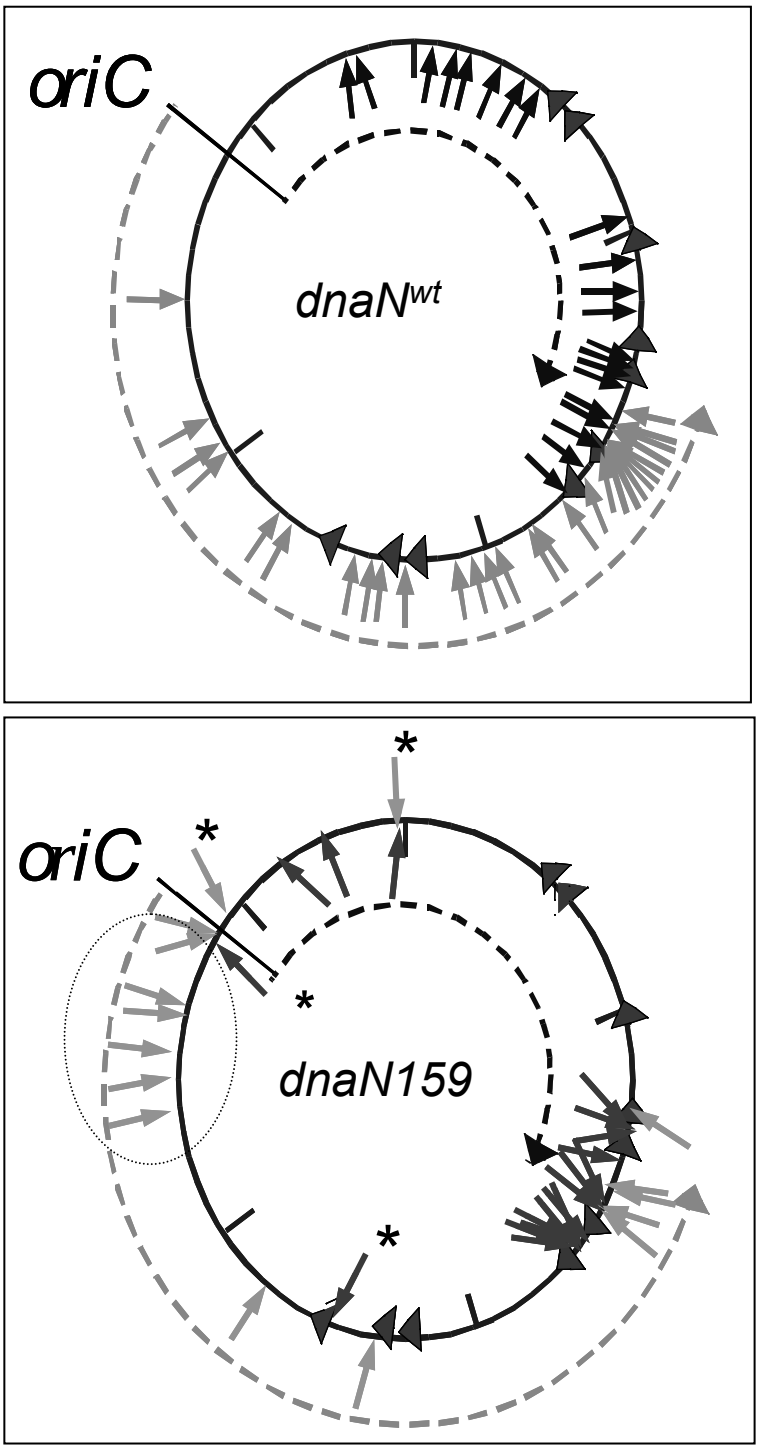
The *dnaN159* allele is highly pleiotropic. Among the phenotypes displayed by this *dnaN* allele are a chronic induction of the SOS response, UV sensitivity, and temperature sensitivity. Mark Sutton and colleagues have proposed that poor interaction of the Pol III α subunit with the β protein produced by the *dnaN159* could partially explain the some of the phenotypes that are observed in this background (Maul et al., 2007; Sutton, 2004). In this model, the affected β clamp would be left behind at a double-stranded single-stranded DNA junction (Heller and Marians, 2005; McInerney and O'Donnell, 2004). Interaction with the clamp on the leading stand would be as affected by the poor interaction as the lagging-strand, possibly resulting in an increase in discontinuous synthesis on the leading-strand as well as the lagging-strand. The leading-strand is thought to be somewhat discontinuous even in wild-type strains, but far less so than the lagging-strand (Heller and Marians, 2005). We decided to map TnsABC+E mediated insertions in the *dnaN159* background to see if we could find evidence of increased availability of β clamps on the leading-strand. We expected that TnsE would be able to direct insertion into the leading-strand, a condition that is not typically seen (Peters and Craig, 2001a), if the leading-strand was replicated in a more discontinuous fashion. We found that a minority of insertions do occur in the leading-strand in the *dnaN159* background, but these insertions are still greatly outnumbered by

insertion into the lagging-strand (Figure 3.9). Interestingly, we also saw an asymmetric shift of the TnsE-mediated insertions on the left side of the chromosome, toward the origin. This may indicate that there are sequences in the left side of the chromosome that slow or impare the progression of DNA replication, possibly highly transcribed ribosomal operons (Figure 3.9). Breier et. al. have pointed out the there is an asymmetry in replication of the *E. coli* chromosome in that the leftward replication starts out faster than the rightward fork (Breier et al., 2005). After a given time, the rightward replication fork catches up and they both meet at the terminus at roughly the same time. TnsE-mediated insertion events also indicate an asymmetry in chromosomal replication that is consistent with that reported by Breier et. al. These insertions may reflect an increased frequency of stalled DNA replication forks, or and increase in the availability of β clamps by some other mechanism, in the left side of the chromosome toward the origin in the *dnaN159* background.

TnsE is capable of disrupting normal coordination of host protein interaction with the processivity factor.

What are the consequences to the host cell when a foreign protein binds to the processivity factor? We reasoned that since TnsE is able to bind the β factor, over-expression of TnsE might interrupt the coordination of host proteins that normally bind to the clamp. To test this idea we monitored SOS induction following over-expression of TnsE. The SOS response is a regulatory network that is normally repressed until DNA damage occurs. Persistently stalled, blocked, or collapsed replication forks are known to trigger the SOS response (Kuzminov, 1995). Constitutive induction of the SOS response is a phenotype associated with some mutant β clamp proteins

Figure 3.9. TnsE-mediated transposition mapped in *dnaN159* background
In the *dnaN159* background, the TnsABC+E insertion profile is shifted toward the origin of replication (gray dashed circle), and the orientation bias appears to be less pronounced (asterisks). Solid circles represent the *E. coli* chromosome. Dashed arrows indicate the direction of replication fork progression. Grey arrows outside the circles and black arrows inside the circles indicate opposite orientations of insertion events collected from independent experiments. The wild-type *dnaN* transposon map has been reproduced from (Peters and Craig, 2001a), and is a combination of TnsEwt and increased-activity mutant TnsE-mediated insertions.


oriC
dnaN^wt
oriC
*
*
*
*
dnaN159

(Maul et al., 2007; Sutton, 2004). We found that over-expression of TnsE does induce the SOS response (Figure 3.10). Moreover, we found that the ability of TnsE to induce SOS is dependent on the putative β clamp interaction motif. Each of the *tnsE*$^{\beta MA}$ mutants significantly reduced ($p<0.0005$, two-tailed unequal variance t-test) the level of SOS induction (Figure 3.10.). Consistent with the yeast two-hybrid results that indicated that the N-terminal 36 amino acids of TnsE are required for interaction with β, expression of proteins that have this region deleted does not induce the SOS response (Figure 3.10). Western blots indicated that TnsE$^{\Delta 1\text{-}36}$ is produced, and appears to be even more stable than wild-type TnsE (data not shown).

Certain mutations in the TnsE protein result in an increase in TnsABC+E transposition frequency (see Chapter 2, Figure 2.8.A). We tested to see if any of these mutations induced the SOS response at expression levels at which wild-type protein does not. We found that the single gain of activity mutation in the N-terminus (M37I) does induce the SOS response when expressed from a *lac* promoter while the others do not (Figure 3.11). TnsE expressed from a *lac* promoter, results in higher protein concentrations than when produced from the endogenous TnsE promoter, but not as high as when produced from the arabinose promoter in pBAD plasmids. The highest increased activity mutant of TnsE is a double mutant (A453, D523N) and is unstable when expressed from the *lac* promoter in plasmids (Joseph Peters, unpublished observation). Under expression of the arabinose promoter, which can be shut off more stringently, does induce the SOS response to a greater extent than wild-type TnsE, but not M37I, following induction of the promoter (Figure 3.12). When the N-terminal 36 amino acids are truncated from this mutant, SOS induction is returned to background levels (Figure 3.12). These

Figure 3.10. Over-expression of TnsE results in activation of the SOS DNA damage response. Activation of SOS is consistent with interference with normal traffic on the β clamp. Cells that over-express wild type TnsE activate the SOS response. Strains expressing the $TnsE^{\Delta 1\text{-}36}$ and $TnsE^{\beta MA}$ mutant proteins show a significant reduction in the SOS response compared to the wild type protein, likely due to a compromised interaction with β. SOS was monitored by Miller assay and in a reporter strain containing a *sulA::lac* fusion and is reported in Miller units(Miller, 1992; Sutton, 2004). Error bars indicate standard error of the mean (n=3). Western blots using antibodies against TnsE are shown bellow each mutant to show similarity in protein expression.

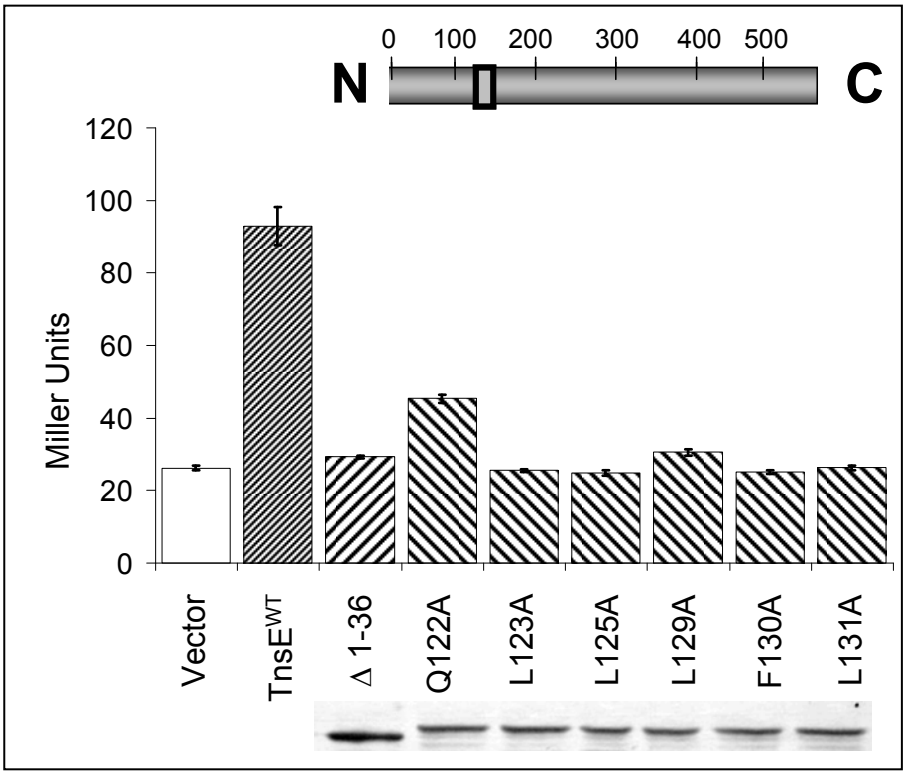

0
100
200
300
400
500
N
C
120
100
80
60
40
20
0
Miller Units
Vector
TnsE^WT
Δ 1-36
Q122A
L123A
L125A
L129A
F130A
L131A

Figure 3.11. The increased transposition activity mutant TnsE^{M37I} mutation induces SOS while other increased activity mutants do not. The C-terminal mutations results in increased DNA binding ability, M37I does not. M37I may constitute a class of mutation affects the ability of TnsE to bind to the β clamp. Error bars indicate standard error of the mean (n=3).

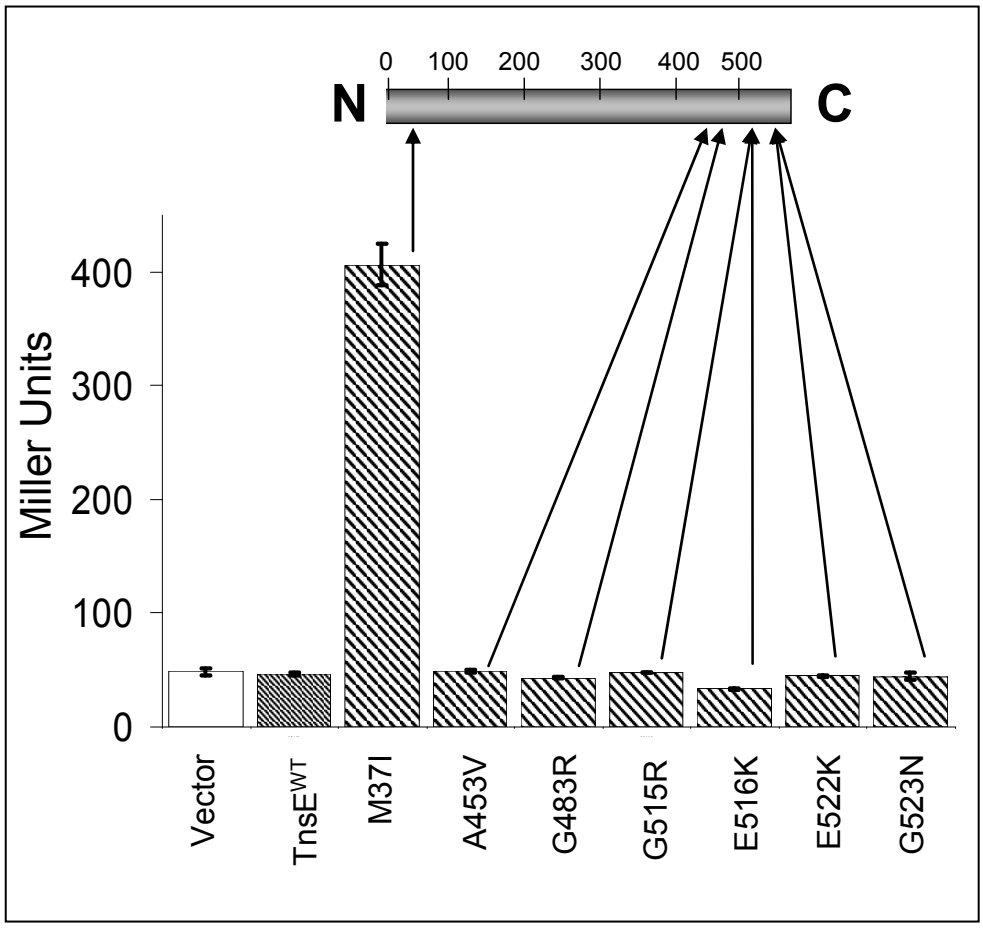
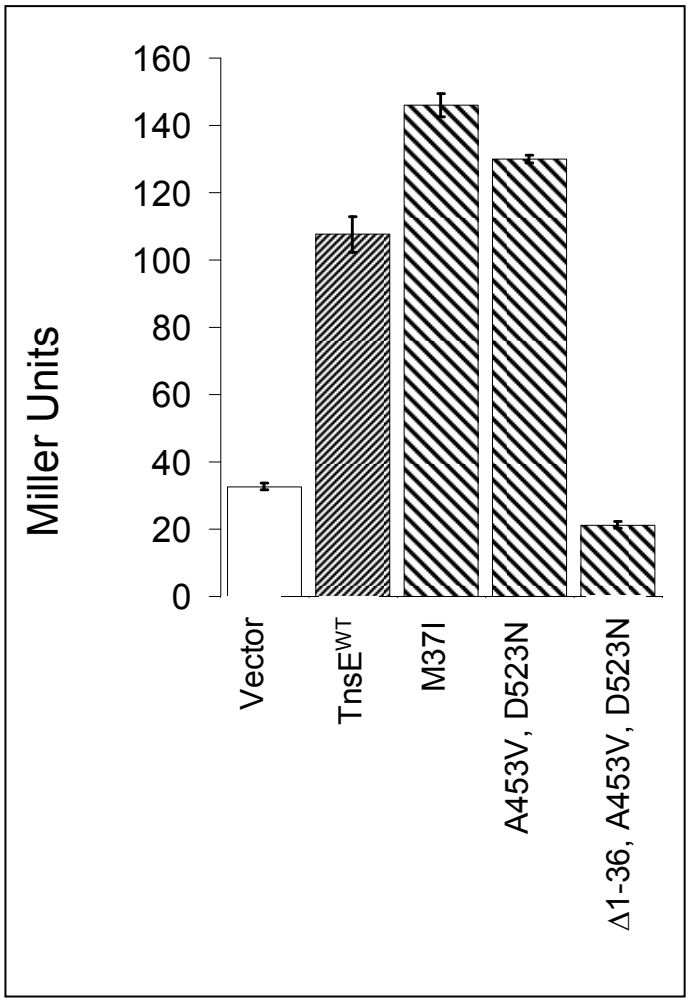

0
100
200
300
400
500
N
C
Miller Units
400
300
200
100
0
Vector
TnsE^WT
M37I
A453V
G483R
G515R
E516K
E522K
G523N

Figure 3.12. Mutant and truncated TnsEs in the SOS induction assay. Over-expression of increased activity mutants $TnsE^{M37I}$ and $TnsE^{A453V,D523N}$ induce SOS to a higher degree than wild-type protein. Deletion of the N-terminal 36 amino acids abrogates SOS induction in $TnsE^{A453V,D523N}$ and $TnsE^{WT}$. Error bars indicate standard error of the mean (n=3).


Miller Units
160
140
120
100
80
60
40
20
0
Vector
TnsE^WT
M37I
A453V, D523N
Δ1-36, A453V, D523N

results suggest that the C-terminus is not likely to be responsible for the SOS induction phenotype, although it may exacerbate the effect in some way. The N-terminus is likely to be instrumental in allowing TnsE to bind to the β clamp, and is essential for the SOS phenotype. The M37I mutation may increase the affinity of TnsE for the β clamp, or it may allow TnsE to access the clamp in ways that wild-type TnsE can not (i.e. bind to other surfaces that are normally occluded by other proteins).

Similar to the constitutive SOS phenotype observed in cells expressing a defective β protein produced by the *dnaN159* allele (Flores et al., 2005), elimination of the RecFOR system suppressed the SOS induction phenotype in cells over-expressing TnsE (Figure 3.13). This result indicates that over-expression of TnsE results in single-stranded DNA gaps, not double-stranded DNA breaks (Kuzminov, 1995). TnsE expressed at wild-type levels does not induce SOS (data not shown), and TnsE-mediated transposition is not dependent on the SOS response since all in vivo transposition assays in this study were carried out in *recA*$^{-}$ cells. These experiments indicate that TnsE has evolved to minimize interference with normal DNA metabolism while maintaining the ability to interact with the essential processivity factor for use in targeting transposition.

We used the SOS induction phenotype as a proxy for testing interaction between β and TnsE proteins from heterologous Tn7-like elements. We expressed TnsE from Tn7-like elements found in *Bacillus cereus ATTC10987* (TnsE^{Bc}), *Shewanella Baltica OS155* (TnsE^{Sb}), and *Idiomarina loihiensis L2TR* (TnsE^{Il}) from arabinose promoters. Under the conditions of the Miller assay (i.e. in liquid LB culture with rapidly growing cells) TnsE^{Bc} is the only TnsE homolog that elicited the SOS response (Figure 3.14). On MacConkey's

Figure 3.13. SOS induction provoked by TnsE over-expression is abolished in the *recF* background. RecF is essential for the RecFOR pathway of RecA loading onto single-stranded DNA gaps. Error bars indicate standard error of the mean (n=3).

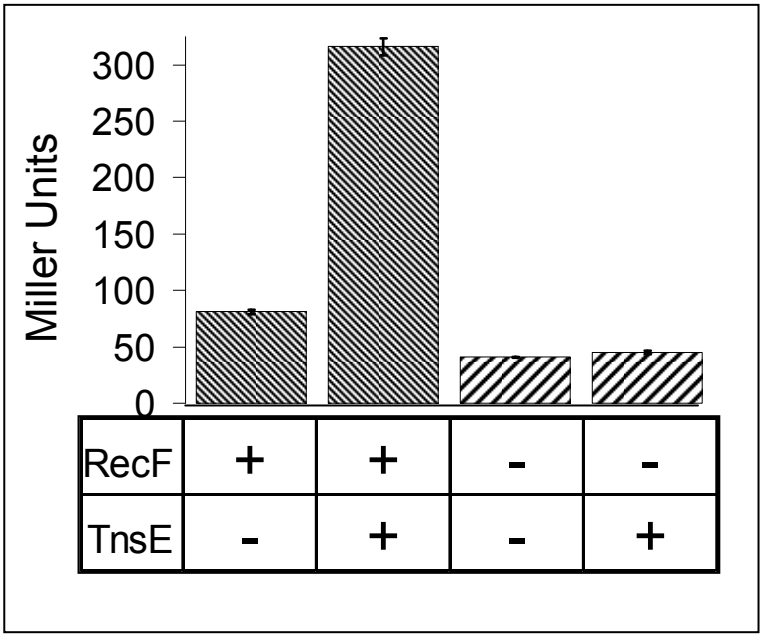
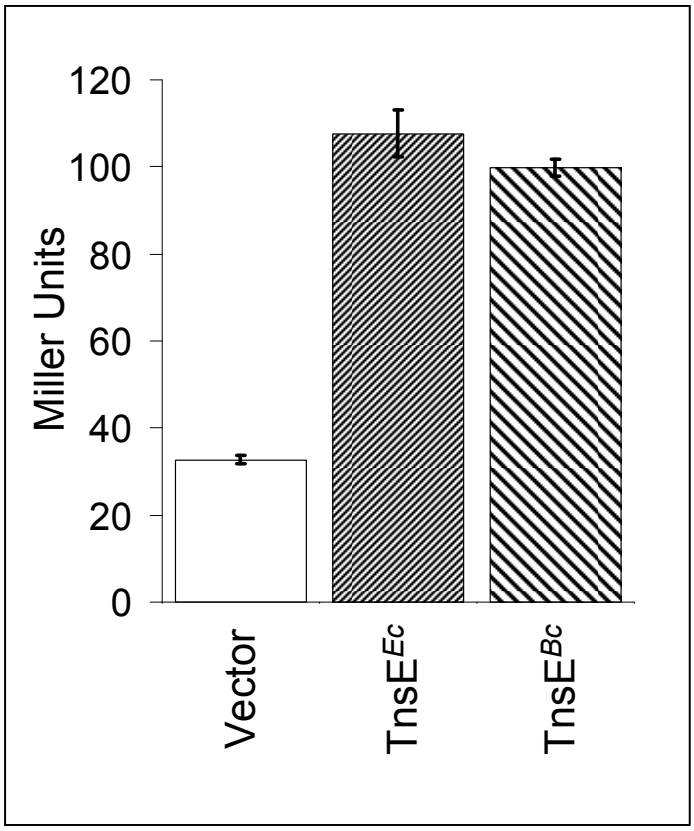

Miller Units
300
250
200
150
100
50
0
RecF + + - -
TnsE - + - +

Figure 3.14. TnsE from *Bacillus cereus* ATCC 10987 also induces the SOS response when controlled by the arabinose promoter. This result provides evidence for an interaction between the β clamp in *E. coli* ($TnsE^{Ec}$) and the TnsE from *B. cereus* ($TnsE^{Bc}$), similar to that seen with TnsE from *E. coli*. Error bars indicate standard error of the mean (n=3).

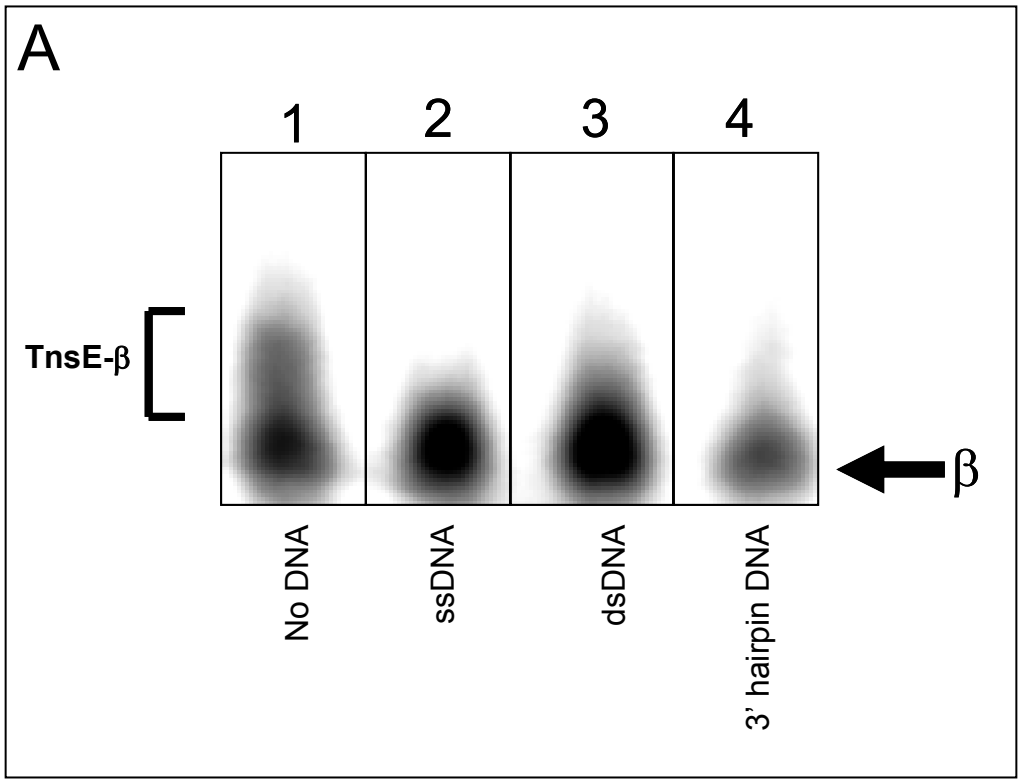

120
100
80
60
40
20
0
Miller Units
Vector
TnsE^Ec
TnsE^Bc

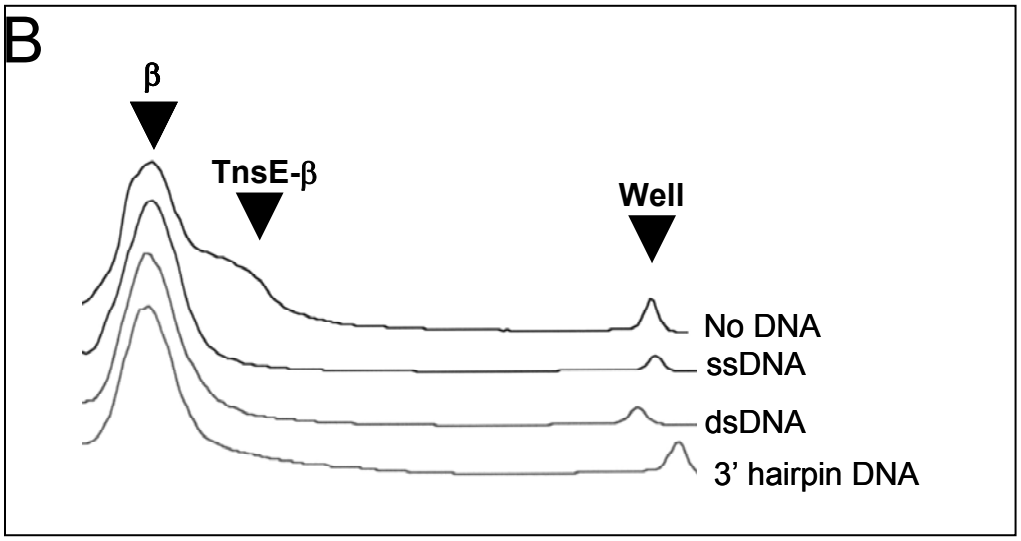
lactose utilization indicator plates supplemented with the same concentration of arabinose as used for the Miller assay, we observed dark red colonies expressing $TnsE^{Sb}$, indicating the induction of the SOS response (data not shown). It is possible that $TnsE^{Sb}$ is only able to disrupt certain kinds of replication, such as repair associated replication that dominates in stationary phase and under the less-than-optimal growth conditions that are present on MacConkey's media.

Interaction with a preferred DNA substrate eliminates interaction with the β clamp.

Since TnsE is a DNA binding protein, we tested to see if interaction with the β clamp is in any way altered when TnsE interacts with its preferred DNA substrate. Previous studies have shown that TnsE binds to many forms of DNA, although it binds preferentially to DNA substrates that present a 3' recessed end. Most proteins that interact with β modulate their activities or β clamp binding affinity based on the DNA substrate present (Johnson and O'Donnell, 2005; Lopez de Saro et al., 2004). With this in mind we decided to test how TnsE interacts with β once bound to DNA. Using the protein-mobility shift assay we assessed the TnsE-β interaction when TnsE had been pre-incubated with excess single-stranded DNA oligos (ssDNA), double-stranded DNA oligos (dsDNA), or a 3' recessed end hairpin structure (3'). We found that, while each of the substrates reduced the mobility shift of ^{32}P-β, the ssDNA and 3' DNA showed no shift in mobility (Figure 3.15). These results suggest that a binding to certain DNA substrates causes TnsE to lose interaction with β in some way. Under the conditions used in the protein-mobility assay, it is highly likely that the ssDNA substrate formed secondary

Figure 3.15. Interaction between TnsE and β is diminished to varying degrees depending on the DNA substrate bound by TnsE.
A. Lane 1 is TnsE and β alone. Lanes 2, 3, and 4 contain a single-stranded DNA oligo, two oligonucleotides annealed to make double-stranded DNA, and an oligo that folds back on itself to make a recessed 3' end structure (respectively). TnsE incubated with single-stranded DNA and the preferred DNA structure, 3' recessed end structures, shows the most marked reduction in β binding. Under the conditions of this experiment, the single-stranded oligonucleotide may form secondary structures that mimic 3' recessed end structures. The bracket indicates shifted products, while the position at which β migrates alone is indicated with an arrow.
B. A graphical representation of the intensity of bands in each lane shows the loss of the majority of TnsE-β complexes in each lane containing each DNA substrate.


A
1
2
3
4
TnsE-β
β
No DNA
ssDNA
dsDNA
3' hairpin DNA

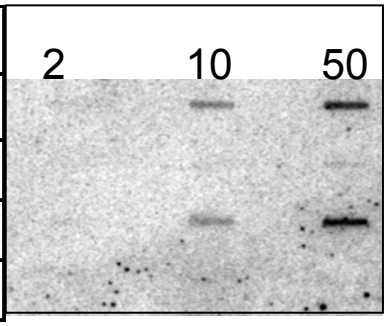

B
β
TnsE-β
Well
No DNA
ssDNA
dsDNA
3' hairpin DNA

structures that might resemble 3' recessed ends and therefore become preferred substrates for TnsE. In addition to this limitation, it has been shown that β itself will interact with ssDNA through the same hydrophobic pocket that is expected to be used by TnsE (Georgescu et al., 2008). We used a separate approach involving far western blots with TnsE cross-linked to various substrates. An excess of single-strand binding protein (SSB) from a hyperthermophillic organism (NEB, ET SSB) was used to prevent secondary structure formation. TnsE, SSB, and the DNA substrates were incubated then cross-linked with formaldehyde and spotted on a membrane in various amounts using a slot-blot apparatus (Experimental procedures). The blots were then probed with ^{32}P-β, and visualized by phosphorimager. We found that, under these conditions, 3'DNA-SSB bound TnsE does not interact with β while ssDNA-SSB incubated TnsE does (Figure 3.16). TnsE incubated with dsDNA-SSB binds weakly to β, while TnsE alone still binds β even after crosslinking. TnsE appears to be a rather flexible protein that may undergo a conformational change upon DNA binding. DNA binding appears to abolish interaction with the β clamp either by occluding the β clamp binding surface itself, or by mediating a conformational change within TnsE that eliminates the β clamp interaction. TnsE may initially interact with β to identify a target molecule, then bind to flanking DNA, releasing the clamp allow space to interact with the transposition machinery. Alternatively, TnsE may bind to the clamp in solution and get loaded onto DNA along with the clamp; there are two binding sites on the clamp that might accommodate interaction with more than one protein (Indiani et al., 2005). These observations suggest a model in which TnsE binds to β in order to identify replicating DNA, then clamp down on the DNA molecule, especially once a 3' end has been found. This may ensure

Figure 3.16. Cross-linked TnsE-DNA structures bind differentially to ^{32}P-β. TnsE, single-stranded binding protein, and various DNA structures were pre-incubated then cross-linked with formaldehyde. Dilutions of these reactions were bound to a nitrocellulose membrane and probed with ^{32}P-β in a far western blot experiment. Row (A) contains TnsE and SSB alone. Row (B) contains TnsE, SSB, and a single-stranded oligonucleotide. Row (C) contains TnsE, SSB, and two annealed oligonucleotides (forming double-stranded DNA). Row (D) contains TnsE, SSB, and a 3' recessed-end-forming hairpin structur.

pmoles complex

	TnsE	SSB	DNA
A	+	+	-
B	+	+	dsDNA
C	+	+	ssDNA
D	+	+	3' hairpin



that transposition directed into double-stranded regions of DNA and not the single-stranded gaps that are expected on the lagging-strand (See discussion, Figure 3.17).

3.5. Discussion

We find that the TnsE protein that is responsible for directing Tn7 transposition into replicating DNA binds to the β processivity factor. A β clamp binding motif found within TnsE and its homologs is required for this interaction in vivo and in vitro (Figure 3.2 and 3.3) and is essential for TnsE activity (Figure 3.6). An increase in the cellular concentration of β leads to an increase in TnsE-mediated transposition (Figure 3.7), and over-expression or mutation of TnsE disrupts normal DNA metabolism (Figure 3.10 and 3.11). We conclude that Tn7 uses the interaction between TnsE and β to direct transposition into discontinuously replicating DNA, especially in mobile plasmids as they enter the cell.

***TnsE utilizes an interaction with the processivity factor to activate and direct transposition into replicating DNA*.**

TnsE-mediated transposition recognizes a facet of lagging-strand DNA replication, directing insertions in a single orientation with this process (Craig, 2002; Peters and Craig, 2001a; Peters and Craig, 2001b). Given that β clamps are found to accumulate on the DNA strand that is replicated discontinuously, the β-TnsE interaction indicates how TnsE-mediated transposition targets the conjugal DNA replication process. Our data suggest that TnsE has evolved to not interrupt the exchange of proteins on clamps during normal DNA replication when expressed at moderate concentrations.

The Kd of the TnsE-β complex is consistent with weak binding in vitro, even without additional host proteins that might exclude TnsE from binding to β. It will be interesting to see how well TnsE competes for access to the clamp in the presence of other proteins that interact with β to carry out normal cellular functions.

We find that we can corrupt the coordination of the replisome by over-expressing TnsE in a process where TnsE proteins capable of interacting with clamps subvert replication, exposing single-strand DNA interruptions that result in induction of the SOS response (Figures 3.10-3.14). The distribution of TnsE-mediated insertions in the chromosome is likely explained by the ability of TnsE to interact with β clamps that are still topologically linked to DNA, but not actively involved in chromosomal DNA replication. For example, in cells lacking mobile plasmids TnsE-mediated insertions occur in the region where DNA replication terminates and sites proximal to repaired DNA double-strand breaks (Peters and Craig, 2000). The SOS induction phenotype that is observed in the various TnsE homologs and mutants appears to be a good indicator of TnsE-β interaction in vivo. With this assay, the only transposon component that is required is the TnsE protein, allowing analysis to be restricted to a single foreign protein without involving all of the transposon machinery. This assay may serve as an extremely informative tool in understanding TnsEs interaction with the host cells replication and repair machinery. Genetic screens involving the SOS induction assay may help reveal new TnsE mutants whose effects specifically result in altered β clamp interaction but not necessarily transposition frequency. Analysis of the transposon targeting profile of these mutants may be useful for understanding exchange of proteins on the β clamp.

The nature of conjugal DNA replication likely leaves β clamps especially vulnerable to TnsE interaction. Conjugal plasmids replicate by a discontinuous process that resembles lagging-strand DNA synthesis that occurs during chromosomal replication. While processivity clamps are expected to be enriched on discontinuously replicating DNA, both on the chromosome and on plasmids, a notable difference is that the discontinuous replication process is not coupled to continuous DNA replication (i.e. leading-strand synthesis) in conjugal DNA synthesis. The protein complex that is present at a standard DNA replication fork may prevent TnsE from binding to the clamp as frequently as during uncoupled DNA replication found in conjugal replication. TnsE-mediated insertion in cells containing the *dnaN159* allele show an altered targeting profile that is consistent with a model in which the main replicative polymerase, PolIII, is impaired for interaction with the clamp and leaves behind clamps on both the leading- and lagging-strands, especially in a particular region in the left side of the chromosome (Figure 3.9).

TnsE appears to lose its ability to interact with the β clamp once it becomes complexed with DNA, especially when the DNA molecule presents a 3' recessed end. This observation leads us to propose a model in which TnsE binds to β clamps that have been abandoned by the replisome, yet still remain topologically linked to DNA. TnsE accomplishes this interaction via the N-terminal domain of the protein. The TnsE may either lose contact with β at this point when it interacts with double-stranded DNA, or TnsE-β complex may then slide along duplex-DNA until it reaches an exposed 3' end, at which point it binds to the DNA and elicits a conformation change that releases the β clamp. Footprinting experiments with TnsD have shown that TnsD introduces a distortion in the DNA that may recruit TnsC in some way

(Bainton et al., 1993; Kuduvalli et al., 2001). Experiments conducted with a mutant TnsABC machinery that does not require TnsE or TnsD for transposition, yet is still sensitive to signals from TnsE and TnsD, suggest that TnsC responds to DNA structure and not necessarily specific protein-protein interactions with target-site selecting proteins (Kuduvalli et al., 2001; Rao and Craig, 2001; Rao et al., 2000). The conformational change upon DNA binding may result a structure that recruits the rest of the transposition machinery, and ultimately leads to the insertion of the Tn7 element (Figure 3.17). By releasing the β clamp, space is made for a nucleoprotein complex can be formed with donor-DNA, target-DNA, and the transposase. The clamp may remain in place and aid in the repair of gaps left by the transposition process at the left end of the newly inserted element. To the host machinery, the gap remaining at the left end may simply appear as an incomplete Okazaki fragment, while the gap on the right end would probably require a separate DNA repair pathway. If this is true, one might expect the left end of the transposon insertion to be repair more efficiently since it would not require the additional step of loading the clamp as would be required for the gap left at the right end of the transposon. Further investigation is required to sort out the details of repair following transposon insertion.

What does Tn7 reveal about coordination of protein traffic on processivity factors?

The processivity factor appears to play a pivotal role in the licensing of activity at the replication fork. TnsE likely binds to the same face of the clamp as MutS, Ligase, Pol III and others (Johnson and O'Donnell, 2005; Lopez de Saro et al., 2006). The presence and importance of β clamp binding motif in

.

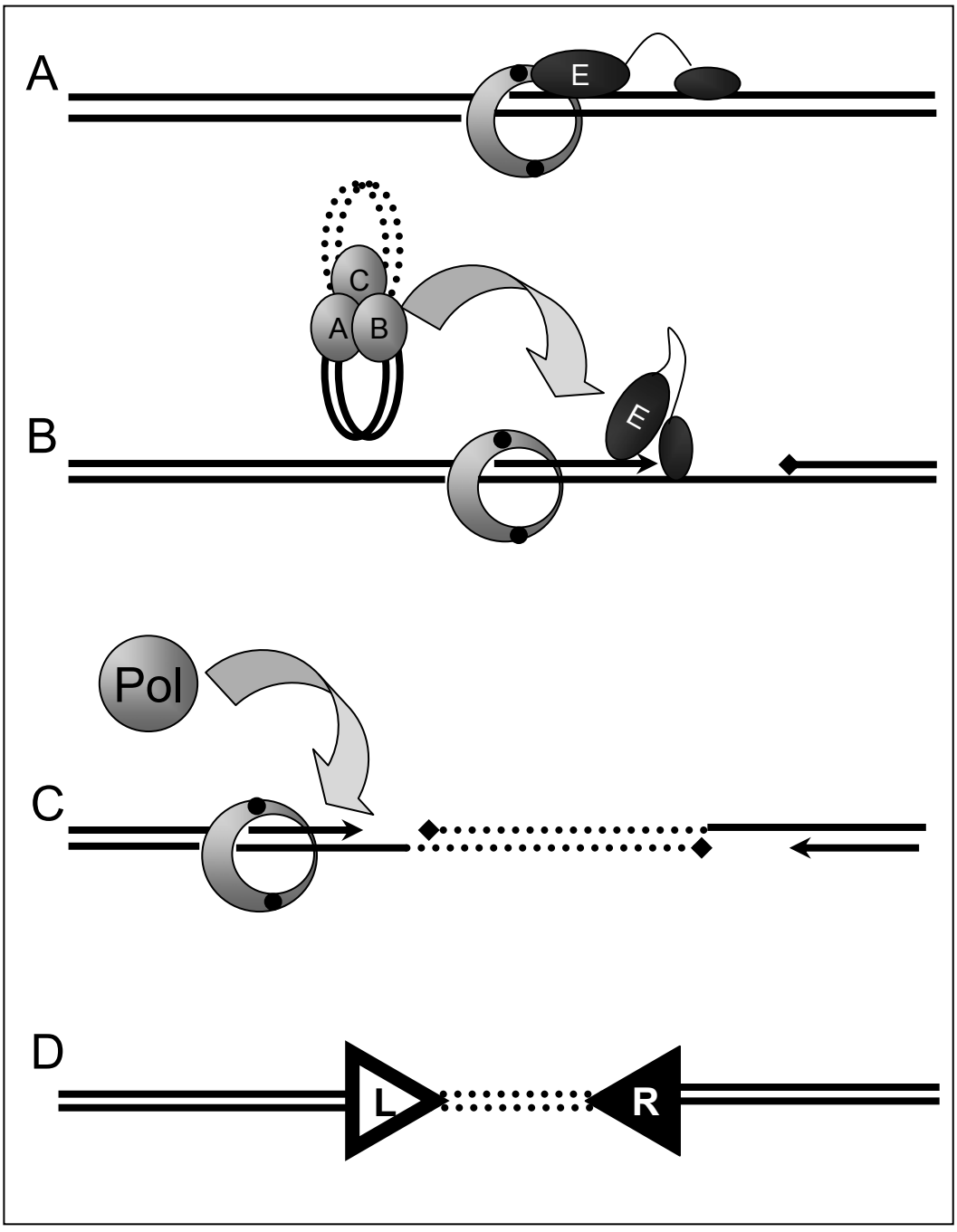
Figure 3.17. Diagram outlining the proposed mechanism of TnsE target identification.

A. TnsE binds to β bound to double-stranded DNA, then scans along DNA for 3' ends. Interaction with the C-terminal face of β, containing the hydrophobic interaction pocket (black circles), orients TnsE in the appropriate direction for scanning the DNA in an efficient way.

B. Once a 3' end has been encountered, The C-terminus of TnsE binds tightly to the DNA structure and undergoes a conformational change, causing the N-terminus to lose contact with β, and possibly distorting the DNA in a manner similar to TnsD. β migrates away from the TnsE-DNA complex to be used later in DNA repair.

C. TnsE recruits the TnsABC core machinery and directs transposition of the Tn7 element (dotted line) proximal to the 3' end structure. The β originally used to identify the target is then used to recruit host machinery (likely PolI) to repair the gap at the left end of the transposon insertion. The gap at the right end would presumably be repair in a separate repair event. Arrow heads represent 3' ends, and filled diamonds represent 5' ends.

D. The resulting left (L) to right (R) orientation of the transposon ends would correspond with the the 5' to 3' direction of the nescient lagging-strand and the polarity of the interacting face of the β clamp.


A
E
C
A
B
E
B
Pol
C
D
L
R

TnsE suggests that interaction with the clamp occurs at least in part through the same hydrophobic pocket on the C-terminal face that appears to be involved in coordination of protein-protein interactions. The SOS induction phenotype observed with TnsE over-expression further supports the notion that TnsE interacts with the clamp on the same face as host proteins involved in DNA replication and repair (Figure 3.10). The SOS induction observed with the M37I mutation suggests that competition for access to β may not be limited to the β clamp binding motif as others have suggested (Lopez de Saro et al., 2004; Wijffels et al., 2004). If SOS induction is interpreted as corruption of normal DNA metabolism due to TnsEs illicit interaction with the clamp, then an increase in the severity of this phenotype might be explained as increased interaction with the clamp, and therefore, enhanced ability to compete for access to the clamp.

The TnsABC+E transposition complex may be using a similar mechanism for detecting strand polarity as has been suggested for mismatch repair systems (MMR) in Eukaryotes and in some Prokaryotes. The ability to interact with only one face of the processivity factor has been suggested to allow strand discrimination in MMR so that newly-replicated DNA containing errors can be selectively removed (Flores-Rozas et al., 2000; Simmons et al., 2008). It stands to reason that an interaction with the β clamp would direct the activity of TnsE in a similarly directional manner, resulting in the orientation bias with replication that we observe with TnsE-mediated insertions. An additional advantage of orienting transposition events with the direction of DNA replication is that after transposition the transcription of transposon genes is aligned with the direction of DNA replication, preventing conflicts between DNA replication and transcription machinery (see Chapter 1).

The TnsE$^{\beta MA}$ mutations did not completely abolish interaction between β and TnsE (Figure 3.16). This result may simply reflect an idiosyncrasy of the far western blot assay, or it may indicate that the TnsE and β interaction occurs over a surface greater than just the β clamp binding motif, as observed with other proteins that interact with β (Bunting et al., 2003; Jeruzalmi et al., 2001). It is thought that the β-clamp binding motif is instrumental in coordinating the order of interaction with the clamp, perhaps by establishing a hierarchy of binding (Lopez de Saro et al., 2004; Wijffels et al., 2004). Interaction with the hydrophobic cleft of the clamp may license further interaction, leading to an active complex between various enzymes and the processivity factor. By mutating individual residues within this interaction motif, the ability of TnsE to compete with other clamp binding proteins may be more severely affected in vivo as compared with the in vitro interaction that is monitored in the absence of all other host proteins. Our findings are consistent with those of other groups who show that proteins that interact with the processivity factor require an intact interaction motif for activity, even though some residual clamp binding is observed in its absence (Beuning et al., 2006; Simmons et al., 2008).

Other transposases bind to the processivity factor and may utilize the interaction for directing transposition into replicating DNA.

Interaction with the processivity factor may constitute a general mechanism for targeting transposition into actively replicating DNA. The transposase of the inactive *pogo* element, found in *Drosophila*, has been shown to bind to the DNA replication processivity factor (PCNA), but the function of this interaction has remained a mystery (Warbrick, 2000; Warbrick

et al., 1998). Proteins found in other inactive transposons that are abundant in humans (*tigger* elements, estimated to be present at ~3000 copies) and in *Arabidopsis* (*lemi1* elements) also possess putative PCNA binding motifs (Warbrick, 2000; Warbrick et al., 1998). Because none of these elements are active, determination of the functional relevance of their interaction with the processivity factor is not possible. We conclude that TnsE has evolved that ability to identify lagging-strand DNA replication through an interaction with the β clamp, likely as a mechanism for targeting processing events found during the mobilization of plasmids. We suggest that a wide range of transposable elements may use a similar mechanism to target DNA replication and/or DNA repair. While mechanistically very different from Tn7 the transposon Tn917 displays target selection profile that resembles that of Tn7 (Garsin et al., 2004). The single polypeptide transposase of Tn917 contains an amino acid sequence (QLCLAR) that resembles the β clamp binding motif described in this work. Tn917-like elements also contain a similar motif and genes mobilized with these elements are causes of antibiotic resistance in medically relevant low G+C Gram positive bacteria (A. Parks and J. Peters, unpublished observation). Additional Eukaryotic transposases may contain PCNA binding motifs (PIP-boxes, consensus sequence QxxLxxFF (Dalrymple et al., 2001)) as well;the transposase of the *Ac* element, found in plants, that has been shown to be stimulated by active DNA replication also contains a PIP-box (QKRIVGFF) (A. Parks and J. Peters, unpublished observation; (Chen et al., 1992)). Interaction with the processivity factor appears to provide transposons with an indication of the replication status of a particular DNA molecule. The presence of a processivity factor indicates the recent passage of a replication fork, and is therefore, an indicator of the presence of multiple copies of that

DNA molecule.

For Tn7, the interaction with the processivity factor appears to be primarily used to activate transposition directed into mobilized plasmids, providing Tn7 with a means of moving to a new host. Since Tn7-like elements are found in a wide variety of hosts, TnsE-mediated transposition shows promise as a new tool for probing the mechanisms and evolution of genetic processes involving processivity factors.

3.6. EXPERIMENTAL PROCEDURES

Plasmids and Strains

All *E. coli* strains were constructed using P1 transduction according to standard genetic techniques (Table 3.1) (Peters, 2007). All primers were purchased from Integrated DNA Technologies (Table 3.2). Strain AP427 was constructed by the method described by Datsenko and Wanner (Datsenko and Wanner, 2000). Briefly, primers JEP191 and JEP192 were used to amplify the Cam^R gene (and flanking *frt* sites) from plasmid pKD3 with 40 bp homology to Kan^R gene cassettes. The PCR product was crossed into AP330 carrying pKD46, which was induced with 0.2% arabinose and the plasmid subsequently lost by growth at 42°C giving AP427.

Plasmid construction was by established methods and detailed in Table 3.3 (Sambrook et al., 1989). Site directed mutagenesis of the putative β clamp binding motif within *tnsE* was carried out using PCR and sub-cloned into the various expression vectors (Table 3.3). The bottom strand primers are shown in Table 3.2 to indicate the specific mutations. Yeast two-hybrid vectors were constructed using the Gateway system (Invitrogen) essentially as described

TABLE 3.1. Strains used in the study

***E. coli* Strains**	**Genotype**	**Reference**
MC4100	*araD139 Δ(argF-lac)169 rpsL150 relA1 flhD5301 deoC1 ptsF25 rbsR22 e14⁻ Δ(fimB-fimE)632::IS1 Δ(fruK-yeiR)725*	(Casadaban, 1976)
NLC28	MC4100 Val^R	(McKown et al., 1988)
NLC51	NLC28 Val^R *recA56*	(McKown et al., 1988)
BB101	B F- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm ΔslyD* (DE3)	(Chivers and Sauer, 1999)
BLR21(DE3)	BL21(DE3) *srl::*Tn*10 ΔrecA*	Mark Sutton
SY2	*Δlac X74 rpsL araD139 Δ(ara-leu)7697 galU galK hsr⁻ hsm⁺sulA::lacZ'YA*::Kan^R	(Ohmori et al., 1995)
JP1386	NLC28 *Δara714*	(Peters and Craig, 2001a)
AP330	JP1386 *sulA::lacZ'YA*::Kan^R	JP1386 X (P1)SY2
AP427	JP1386 *sulA::lacZ'YA*::*frt*	This work
STL3607	AB1157 *recF400*::Kan^R	(Lovett et al., 2002)
AP461	AP427 *recF400*::Kan^R	AP 427 X (P1) STL3607
***Yeast* Strains**	**Genotype**	**Reference**
EGY40 [pSH18-34]	MATa ura3-52 trp1-1 leu2-3,112 [pSH18-34]	(Liachko and Tye, 2005)

TABLE 3.2. Primers used in this study

Primer name	Primer sequence	Use
JEP128	5' – TTC TTT ATC GCC AAG CGT CT – 3'	Amplicon for cloning β and β specific promoters
JEP129	5' – CGG GTG AGG GAC ATT ACA GT – 3'	Amplicon for cloning β and β specific promoters
JEP149	5' – TTA ATT AAA AAT AAT GAC CGA GCT AGC TCT AAC GCT GGG – 3'	Introducing the Q122A mutation into TnsE
JEP150	5' – TTA ATT AAA AAT AAT GAC CGA GCT AGC TCT GCC TGT GGG – 3'	Introducing the L123A mutation into TnsE
JEP151	5' – TTA ATT AAA AAT AAT GAC CGA GCT GCC TCT AAC TGT GGG – 3'	Introducing the L125A mutation into TnsE
JEP152	5' – TTA ATT AAA AAT GCT GAC CGA GCT AGC TCT AAC TGT GGG – 3'	Introducing the L129A mutation into TnsE
JEP153	5' – TTA ATT AAA GCT AAT GAC CGA GCT AGC TCT AAC TGT GGG – 3'	Introducing the F130A mutation into TnsE
JEP155	5' – CAG AAA TAG GAG TTA ATT GCA AAT AAT GAC CGA GCT AG – 3'	Introducing the L131A mutation into TnsE
JEP191	5'- AGA AAA ACT CAT CGA GCA TCA AAT GAA ATGAAA CTG CAA TTT ATT CAT A GT GTA GGC TGG AGC TGC TTC-3'	Amplicon for replacing the Kan^R cassette with a *frt*-Cam^R-*frt* cassette
JEP192	5'-ATG AGC CAT ATT CAA CGG GAA ACG TCT TGC TCG AGG CCGCGA TTA AAT TCA TAT GAA TAT CCT CCT TAG -3'	Amplicon for replacing the Kan^R cassette with a *frt*-Cam^R-*frt* cassette
JEP272	5'- GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GGA GAT AGA ACC ATG GTT AGG CTA GCT ACA TTT AAT GAC – 3'	TnsE fusions for use in the two hybrid assay
JEP273	5'- GGGG AC CAC TTT GTA CAA GAA AGC TGG GTT TTA ATG CGT AAA TTG CTC TC – 3'	TnsE fusions for use in the two hybrid assay

TABLE 3.2 (Continued)

Primer name	Primer sequence	Use
JEP275	5'- GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GGA GAT AGA ACC ATG AAA TTT ACC GTA GAA CG – 3'	β fusions for use in the two hybrid assay
JEP276	5'- GGGG AC CAC TTT GTA CAA GAA AGC TGG GTT TTA CAG TCT CAT TGG CAT GAC – 3'	β fusions for use in the two hybrid assay
JEP278	5' - GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC GAA GGA GAT AGA ACC ATG AAT CGG TTG TAC CCG GAA C – 3'	δ fusions for use in the two hybrid assay
JEP279	5' - GGGG AC CAC TTT GTA CAA GAA AGC TGG GTT TCA ACC GTC GAT AAA TAC GTC – 3'	δ fusions for use in the two hybrid assay
JEP165	5' – TGTTGTTGCTGCCAATCTGTA CGC – 3'	For making ssDNA and dsDNA structures
JEP166	5' – GCGTACAGATTGGCAGCAACA ACA – 3'	Complementary to JEP166 for making dsDNA substrates
JEP167	5' - CCA TTA GCA AGG CCG GAA ACG TCA CCA ATG CAA CGA TCA GCC AAC TAA ACT AGG ACA TCT – 3'	3' recessed end hairpin structure

TABLE 3.3. Plasmids used in this study

Name	Relevant information
pBTM^{gw}	pBTM16 derivative containing a gateway cassette, 2μ, LEU2, GAD4-AD (Liachko and Tye, 2005)
pGAD^{gw}	pGAD2F derivative containing a gateway cassette, 2μ, TRP1, LEXA-DBD (Liachko and Tye, 2005)
pBAD24	pBR322 derivative containing a multi-cloning site under the control of an arabinose promoter, ampicillin resistant cloning vector (Guzman et al., 1995)
pTA106	pSC101 replicon, ampicillin resistant cloning vector
pβHK	a pET16b (Novagen) derivative containing a protein kinase motif and His_6 tag fused to the N-terminus of β, ampicillin resistant (Kelman et al., 1995b)
pCP20	Temperature sensitive plasmid with thermal induction of FLP recombinase (Cherepanov and Wackernagel, 1995)
pKD46	Temperature sensitive plasmid with the λ Red proteins under arabinose control (Datsenko and Wanner, 2000)
pKD3	Plasmid encoding ampicillin resistance that allows PCR amplification of a gene cassette encoding chloramphenicol resistance flanked by *frt* sites recognized by the FLP recombinase (Datsenko and Wanner, 2000)
pCAW11	pET22b (Novagen) derivative encoding TnsE, ampicillin resistant (Wolkow et al., 1996)
pCW15	pACYC184 derivative encoding TnsABC, chloramphenicol resistant (Waddell and Craig, 1988)(For data shown in Figure 3.6)
pJP104	pTA106 derivative encoding TnsE, ampicillin resistant (Peters and Craig, 2001a)(For data shown in Figure 3.6)
pJP131	pBAD24 derivative encoding TnsE, ampicillin resistant (Peters and Craig, 2001a)(For data shown in Figure 3.10, 3.12, 3.13, 3.14)
pMG1	pBAD24 derivative encoding TnsE from *Bacillus cereus* ATCC 10987 (Galac and Peters, unpublished work)(for sata shown in Figure 3.14)
pMG2	pBAD24 derivative encoding TnsE from *Idiomarina loihiensis L2TR* (Galac and Peters, unpublished work)(for sata shown in Figure 3.14)
pMG3	pBAD24 derivative encoding TnsE from *Shewanella baltica OS155* (Galac and Peters, unpublished work)(for sata shown in Figure 3.14)
pQS100	pTA106 vector encoding TnsABC (For data shown in Figure 3.7 and 3.8) (Shi, 2008)
pQS102	pTA106 vector encoding TnsABC+E (For data shown in Figure 3.7 and 3.8)(Shi, 2008)
pARP30	Plasmid pGAD with *tnsE* fused to the yeast transcription activation domain. (For data shown in Figure 3.2)

TABLE 3.3 (Continued)

Name	Relevant information
pARP35	Plamid pBTM with β fused to the GAL4 binding domain. (For data shown in Figure 3.2)
pAP7	Plasmid pGEM-T containing the β gene and all known β specific promoters from the PCR product produced by JEP128 and JEP129 cloned into the T-A cloning site. (For data shown in Figure 3.7)
pARP63	pACYC184 β gene with all known β specific promoters (Armengod et al., 1988) constructed by cloning an *Eag*I fragment from pAP7 into the *Eag*I site within the Tet^R gene of pACYC184. The direction of β transcription is the same as that of the Tet^R gene. (For data shown in Figure 3.7)

in the manufacturers recommendations and elsewhere (Liachko and Tye, 2005). Primers JEP272 and JEP273 were used to amplify *tnsE* using pJP131 as a template (Tables 3.2 and Table 3.3). Primers JEP275 and JEP276 were used to amplify *dnaN* (β gene) in a single colony PCR from strain NLC28 (Tables 1 and S1). Primers JEP278 and JEP279 were used to amplify *holA* (δ gene) in a single colony PCR from strain NLC28 (Tables 3.1 and 3.3). PCR products were gel purified and transferred into pDONOR221 via the Gateway BP reaction as recommended by the vendor. pDONOR221 constructs were confirmed by sequencing. The *tnsE* gene was then transferred to $pGAD^{gw}$ and *dnaN* and *holA* were transferred to $pBTM^{gw}$ via the Gateway LR reaction as recommended by the vendor. $TnsE^{\beta MA}$ mutants were cloned from pJP131 containing $tnsE^{\beta MA}$ mutants into pDONOR221 containing *tnsE*, then transferred via LR reaction to $pGAD^{gw}$ (Liachko and Tye, 2005).

Computational analysis

The *tnsE* genes identified in previous work were aligned using the ClustalW algorithm (Thompson et al., 1994) through the online Jalview server (Clamp et al., 2004)(Table 3.4). A consensus sequence was generated using the Jalview software (Clamp et al., 2004).

Yeast two-hybrid assay

Yeast strain EGY40[pSH18-34] (Golemis, 2002) was co-transformed with $pBTM^{gw}$ and $pGAD^{gw}$ derivatives and selected on SC-Leu-Trp plates. Overnight yeast cultures of each isolate were grown in 5 ml SC–Trp–Leu dropout media (Invitrogen) to an OD_{600} ~1. 1ml of cells were spun down and

TABLE 3.4. Sequences used in computational analysis (Parks and Peters, 2007).

Host Species	Accession #
E.coli	CAA35687
P.stuartii	ABG21684
S. putrefaciens 200	EAY53112
I. loihiensis L2TR	AA83441
S. baltica OS155	ABN63817
S. putrefaciens CN-32.2	ABP77648
S. putrefaciens CN-32.1	ZP_00814471
S. loihica PV-4	ABO25690
P. carbinolicus	YP_358326
H. chejuensis KCTC2396	YP_438100
A. ferrooxidans	AAC21663
B. cereus ATCC10987	AAS39122
Staphylococcus sp.693-2	ABG49263
Immobile Ac/T-DNA vector pNU400	ABB59986.1
Tn917 transposase sequence	AF061336

resuspended in Z-buffer. Miller assays were performed as described by Amberg et. al. (Amberg et al., 2006).

Protein purification and labeling

His-6-tagged TnsE proteins were purified as described (Peters and Craig, 2001a). A modified β protein that could be labeled with ^{32}P phosphate was purified using the previously described vector (Kelman et al., 1995b) according to a previously method (Turner et al., 1999). β was labeled with ^{32}P as described using αAMP Protein Kinase (PKA) purchased from NEB (Kelman et al., 1995a).

Protein Gel Mobility Shift Assays

Native polyacrylamide gel electrophoresis assays using ^{32}P-β were performed as described (Lopez de Saro and O'Donnell, 2001). Each 15 ul reaction was composed of 25 mM Tris-Cl (pH 7.5), 0.5 mM EDTA, 10% glycerol, 50 μg/ml bovine serum albumin, 100 mM KCl, 5 mM DTT, and 50 nM ^{32}P-β. Samples were incubated at 37 °C for 5 min, then 10 μl of the reaction was run on a 4% native polyacrylamide gel (4% acrylamide:bisacrylamide 29:1, 1XTBE buffer, 5% glycerol). Electrophoresis was performed in 1XTBE buffer (90 mM Tris borate, 65 mM boric acid, 2.5 mM EDTA) at 17mA for two hours (23°C). Gels soaked in 20% ethanol and 5% glycerol for 10 minutes then dried overnight. ^{32}P-β was detected using a PhosphorImager (Molecular Dynamics) and quantified using ImageQuant software (Amersham Biosciences).

Far Western Blot

An Immobilon-P (Millipore) membrane was prepared according to manufacturers recommendations. Proteins were spotted on the wet membrane in the quantities noted in the legend. The membrane was then blocked overnight at 25°C, in blocking buffer containing 20 mM HEPES (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1% Triton X-100, 1% BSA. The membrane was probed as described previously (Einarson et al., 2007). An interaction buffer composed of 20 mM HEPES (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1% Triton X-100, 5% glycerol, 1% BSA, and containing 5 nM ^{32}P-β was used to probe the membrane at 4°C for 3.5 hours. Following interaction, the membrane was washed four times with PBS (10 mM Na-phosphate, pH7.2, 0.9% NaCl, 0.2% Triton X-100), and twice with PBS augmented with 100 mM KCl. The membrane was dried for 30 min at 42°C and ^{32}P-β was detected using a PhosphorImager (Molecular Dynamics) and quantified using ImageQuant software (Amersham Biosciences).

Equilibrium gel filtration

A 10ml column (250 mm X 7.5 mm) was packed with 4% agarose beads (MP Biomedicals) and equilibrated with buffer containing 10 mM Tris-HCl (pH7.6) and 1nM ^{32}P-β. 790 pmoles (or 200 pmoles) of TnsE was loaded onto the column in the same buffer containing 1nM ^{32}P-β (or 50 nM). 145ul fractions were eluted from the column, and 50ul of each fraction was analyzed by liquid scintillation counter. Radiation signal was converted to fmoles of ^{32}P-labeled β and plotted vs. total volume eluted. Peaks indicate TnsE-β complexes, while the trough is evidence of depletion of ^{32}P-β due to complex formation (Beeckmans, 1999). The amount of additional ^{32}P-β eluted in the

peaks was calculated by determining the area under the curve. The Kd of the TnsE-β complex was calculated by determining using the equation Kd=([free β][free TnsE]/[TnsE-β complex]) = (50*(703.7/86.3)) = 408 nM (Beeckmans, 1999).

Transposition Assay

Transposition assays were conducted in the strain NLC51 containing *tns* genes encoded on plasmids as described (McKown et al., 1988). A miniTn7 element containing a Kam^R cassette was introduced using a defective lambda phage (λKK1) that cannot integrate or replication in the NLC51 background strains used in this assay. Transposition frequency was obtained by dividing the number of Kanamycin resistant colonies by the number of infectious λKK1 phage used in the assay.

SOS induction assay

AP330 or its derivatives containing pBAD24 derivatives were grown overnight in LB media containing 0.2% glucose and ampicillin (100 μg/ml) at 37°C. Overnight cultures were diluted 1:100 in fresh LB with 0.2% arabinose and ampicillin then grown two hours at 37°C. β galactosidase activity was measured by the Miller assay as described (Miller, 1992).

3.7. ACKNOWLEDGEMENTS

This work was supported by NIH grant R01 GM069508. I thank Mark Sutton and Bik Tye for strains and plasmids and helpful discussions. I also thank Eric Alani, Bik Tye, Jeff Roberts, and the members of the Peters laboratory for comments on early drafts of this Chapter. I am greatful to Madeline Galac for the production of the plasmids containing TnsE homologs from *Idiomarina loihiensis* L2TR, *Shewanella baltica* OS155, and *Bacillus cereus* ATCC10987.

REFERENCES

Amberg, D. C., Burke, D. J., and Strathern, J. N. (2006). Assay of {beta}-Galactosidase in Yeast: Permeabilized Cell Assay 10.1101/pdb.prot4158. Cold Spring Harbor Protocols *2006*, pdb.prot4158-.

Armengod, M. E., Garcia-Sogo, M., and Lambies, E. (1988). Transcriptional organization of the dnaN and recF genes of *Escherichia coli K-12*. J Biol Chem *263*, 12109-12114.

Bainton, R. J., Kubo, K. M., Feng, J.-N., and Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell *72*, 931-943.

Becherel, O. J., Fuchs, R. P., and Wagner, J. (2002). Pivotal role of the beta-clamp in translesion DNA synthesis and mutagenesis in *E. coli* cells. DNA Repair (Amst) *1*, 703-708.

Beeckmans, S. (1999). Chromatographic methods to study protein-protein interactions. Methods *19*, 278-305.

Beuning, P. J., Sawicka, D., Barsky, D., and Walker, G. C. (2006). Two processivity clamp interactions differentially alter the dual activities of UmuC. Mol Microbiol *59*, 460-474.

Breier, A. M., Weier, H. U., and Cozzarelli, N. R. (2005). Independence of replisomes in *Escherichia coli* chromosomal replication. Proc Natl Acad Sci U S A *102*, 3942-3947.

Bunting, K. A., Roe, S. M., and Pearl, L. H. (2003). Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. Embo J *22*, 5883-5892.

Chen, J., Greenblatt, I. M., and Dellaporta, S. L. (1992). Molecular analysis of Ac transposition and DNA replication. Genetics *130*, 665-676.

Cherepanov, P. P., and Wackernagel, W. (1995). Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9-14.

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java alignment editor. Bioinformatics *20*, 426-427.

Craig, N. L. (2002). Tn7. In Mobile DNA II, M. Lambowitz-Alan, ed. (Washington, DC, ASM Press), pp. 423-456.

Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E., and Jennings, P. A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proc Natl Acad Sci U S A *98*, 11627-11632

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

Einarson, M. B., Pugacheva, E. N., and Orlinick, J. R. (2007). Far Western: Probing Membranes 10.1101/pdb.prot4759. Cold Spring Harbor Protocols *2007*, pdb.prot4759-.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Flores, M. J., Sanchez, N., and Michel, B. (2005). A fork-clearing role for UvrD. Mol Microbiol *57*, 1664-1675.

Flores-Rozas, H., Clark, D., and Kolodner, R. D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. Nat Genet *26*, 375-378.

Garsin, D. A., Urbach, J., Huguet-Tapia, J. C., Peters, J. E., and Ausubel, F. M. (2004). Construction of a *Enterococcus faecalis* Tn917-mediated-gene-disruption library offers insight into Tn917 insertion patterns. Journal of Bacteriology *186*, 7280-7289.

Georgescu, R. E., Kim, S. S., Yurieva, O., Kuriyan, J., Kong, X. P., and O'Donnell, M. (2008). Structure of a sliding clamp on DNA. Cell *132*, 43-54.

Golemis, E. (2002). Protein-Protein Interactions (Cold Spring Harbor NY, CSHL Press).

Guzman, L.-M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and highlevel expression by vectors containing the arabinose P_{BAD} promoter. J Bact *177*, 4121-4130.

Heller, R. C., and Marians, K. J. (2005). The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. Mol Cell *17*, 733-743.

Indiani, C., McInerney, P., Georgescu, R., Goodman, M. F., and O'Donnell, M. (2005). A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. Mol Cell *19*, 805-815.

Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., and Kuriyan, J. (2001). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. Cell *106*, 417-428.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem *74*, 283-315.

Kelman, Z., Naktinis, V., and O'Donnell, M. (1995a). Radiolabeling of proteins for biochemical studies. Methods Enzymol *262*, 430-442.

Kelman, Z., Yao, N., and O'Donnell, M. (1995b). *Escherichia coli* expression vectors containing a protein kinase recognition motif, His6-tag and hemagglutinin epitope. Gene *166*, 177-178.

Kuduvalli, P., Rao, J. E., and Craig, N. L. (2001). Target DNA structure plays a critical role in Tn7 transposition. EMBO J *20*, 924-932.

Kuzminov, A. (1995). Instability of inhibited replication forks in *E. coli*. Bioessays *17*, 733-741.

Lenne-Samuel, N., Wagner, J., Etienne, H., and Fuchs, R. P. (2002). The processivity factor beta controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis in vivo. EMBO Rep *3*, 45-49.

Liachko, I., and Tye, B. K. (2005). Mcm10 is required for the maintenance of transcriptional silencing in Saccharomyces cerevisiae. Genetics *171*, 503-515.

Lopez de Saro, F., Georgescu, R. E., Leu, F., and O'Donnell, M. (2004). Protein trafficking on sliding clamps. Philos Trans R Soc Lond B Biol Sci *359*, 25-30.

Lopez de Saro, F. J., Marinus, M. G., Modrich, P., and O'Donnell, M. (2006). The beta sliding clamp binds to multiple sites within MutL and MutS. J Biol Chem *281*, 14340-14349.

Lopez de Saro, F. J., and O'Donnell, M. (2001). Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I. Proc Natl Acad Sci U S A *98*, 8376-8380.

Maul, R. W., Ponticelli, S. K., Duzen, J. M., and Sutton, M. D. (2007). Differential binding of Escherichia coli DNA polymerases to the beta-sliding clamp. Mol Microbiol *65*, 811-827.

McInerney, P., and O'Donnell, M. (2004). Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. J Biol Chem *279*, 21543-21551.

McKown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988). Sequence requirements of *Escherichia coli attTn7*, a specific site of transposon Tn7 insertion. J Bacteriol *170*, 352-358.

Miller, J. H. (1992). A short course in bacterial genetics (New York, Cold Spring Harbor).

Peters, J. E. (2007). Gene transfer - Gram-negative bacteria, Chapter 31. In Methods for General and Molecular Microbiology, S. L. R., ed. (Washington D. C., ASM Press), pp. 735-755.

Peters, J. E., and Craig, N. L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol Cell *6*, 573-582.

Peters, J. E., and Craig, N. L. (2001a). Tn7 recognizes target structures associated with DNA replication using the DNA binding protein TnsE. Genes & Dev *15*, 737-747.

Peters, J. E., and Craig, N. L. (2001b). Tn7: smarter than we thought. Nature Reviews/Molecular Cell Biology *2*, 806-814.

Rao, J. E., and Craig, N. L. (2001). Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon Tn7. J Mol Biol *307*, 1161-1170.

Rao, J. E., Miller, P. S., and Craig, N. L. (2000). Recognition of triple-helical DNA structures by transposon Tn7. Proc Natl Acad Sci USA *97*, 3936-3941.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd edn (Cold Spring Harbor, NY, Cold Spring Harbor Press).

Sharpe, P., and Craig, N. L. (1998). Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. EMBO J *17*, 5822-5831.

Shi, Q., Parks, A.R., Potter, B.D., Safir, I.J., Luo, Y., Forster, B.M., Peters, J.E. (2008). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in E. coli. Genetics *In Press*.

Simmons, L. A., Davies, B. W., Grossman, A. D., and Walker, G. C. (2008). Beta clamp directs localization of mismatch repair in *Bacillus subtilis*. Mol Cell *29*, 291-301.

Sutton, M. D. (2004). The *Escherichia coli dnaN159* mutant displays altered DNA polymerase usage and chronic SOS induction. J Bacteriol *186*, 6738-6748.

Sutton, M. D., and Duzen, J. M. (2006). Specific amino acid residues in the beta sliding clamp establish a DNA polymerase usage hierarchy in *Escherichia coli.* DNA Repair (Amst) *5*, 312-323.

Sutton, M. D., Duzen, J. M., and Maul, R. W. (2005). Mutant forms of the Escherichia colibeta sliding clamp that distinguish between its roles in replication and DNA polymerase V-dependent translesion DNA synthesis. Mol Microbiol *55*, 1751-1766.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Turner, J., Hingorani, M. M., Kelman, Z., and O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. Embo J *18*, 771-783.

Waddell, C. S., and Craig, N. L. (1988). Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes & Dev *2*, 137-149.

Warbrick, E. (2000). The puzzle of PCNA's many partners. Bioessays *22*, 997-1006.

Warbrick, E., Heatherington, W., Lane, D. P., and Glover, D. M. (1998). PCNA binding proteins in *Drosophila melanogaste*r : the analysis of a conserved PCNA binding domain. Nucleic Acids Res *26*, 3925-3932.

Wijffels, G., Dalrymple, B. P., Prosselkov, P., Kongsuwan, K., Epa, V. C., Lilley, P. E., Jergic, S., Buchardt, J., Brown, S. E., Alewood, P. F.*, et al.* (2004). Inhibition of protein interactions with the beta 2 sliding clamp of *Escherichia coli* DNA polymerase III by peptides from beta 2-binding proteins. Biochemistry *43*, 5661-5671.

Wilkins, B., and Lanka, E. (1993). DNA processing and replication during plasmid transfer between gram-negative bacteria. In Bacterial Conjugation, D. B. Clewell, ed. (New York, Plenum Press).

Wolkow, C. A., DeBoy, R. T., and Craig, N. L. (1996). Conjugating plasmids are preferred targets for Tn7. Genes & Dev *10*, 2145-2157.

CHAPTER 4. GENETIC ANALYSIS OF TNSE-MEDIATED TRANSPOSITION INTO THE *E. coli* CHROMOSOME[4]

4.1 Summary

The transposon Tn7 has the ability to specifically recognize two distinct types of targets in host organisms. The TnsE pathway most likely evolved to direct transposition into mobile DNA targets. At a lower frequency, TnsE-mediated transposition events occur in the bacterial chromosome. Certain mutations in the host machinery affect the frequency of transposition into the chromosome. Analysis of transposition in the context of these mutations should provide insight into the molecular targets that are identified by TnsE on the chromosome and improve understanding of the events that occur during genome maintenance. Here I present analysis of selected host cell mutations that affect or might be expected to affect TnsE-mediated transposition. I find that TnsE does not rely on any specific homologous recombination or replication restart pathway for identification of transposition targets on the chromosome. TnsE-mediated transposition is stimulated in certain mutant backgrounds, hinting at the use of alternate genome maintenance pathways

[4] Parts of this chapter appear in a paper published in the journal Genetics Shi, Q., Parks, A.R., Potter, B.D., Safir, I.J., Luo, Y., Forster, B.M., Peters, J.E. (2008). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in E. coli. Genetics *In Press*. All of the work presented in this chapter was conducted by A.R.P. Figures 4.2 and 4.3 are published in Shi, et. al. 2008.

that may present more frequent or prolonged appearance of structures or complexes recognized by TnsE.

4.2 Introduction

The TnsABC+E pathway of transposition most likely evolved as a means of selecting Tn7 transposition targets that are actively transported between cells. This assumption is based on the presumed evolutionary advantage of preferentially directing transposition into mobile DNA, and based on the higher frequency of transposition into conjugal plasmids compared to the frequency of transposition into the chromosome. It appears that TnsE identifies these targets through an interaction with the β processivity factor (see Chapter 3). Utilization of such a ubiquitous landmark for the identification of targets also results in activation of transposition into the chromosome, a process that is thought to be an artifact of this target-site selection pathway. Insertion into the chromosome via the TnsE pathway occurs at a much lower frequency than transposition into mobile plasmids. These insertions are enriched in the region of the chromosome where DNA replication terminates, and occur with orientation of the transposon ends in a way that is consistent with the recognition of lagging-strand DNA synthesis. In addition to targets in the terminus region, TnsABC+E transposition occurs proximal to DNA double-strand breaks (DSBs) that are induced at specific locations in the chromosome. Moreover, these insertions occur with a specific orientation in discrete “hotspots” that are activated based on the location of the induced DSBs (Shi, 2008).

In Chapter 3, I proposed that target-site selection in the TnsE pathway is accomplished through a specific interaction with the β processivity factor,

and that the location and orientation of transposon insertions can be explained by enrichment and orientation of β clamps in those chromosomal locations. Given the observations presented in Chapter 3, one might suggest that complexes that are present on the chromosome might resemble conditions found during conjugal replication and "fool" TnsE into activating transposition into chromosomal sites where those conditions exist. In addition to binding to the β clamp there may be specific conditions, such as the presence of a given DNA structure or other host factors, which are required to activate transposition via the TnsE pathway. Alternatively, there may be particular circumstances under which TnsE is allowed to interact with the β clamp or an enrichment of β clamps may be required to activate transposition. In this chapter, I examine TnsE-mediated insertions in the absence of conjugal replication, and with the deletion or mutation of selected host genes to provide insight into the requirements for TnsE-mediated transposition into the chromosome. This type of analysis may also refine our understanding of the molecular structures that are present during the maintenance of the chromosome by host machinery.

The modular nature of Tn7 makes it a useful and tractable genetic tool for analyzing genetic phenomena. Since Tn7 can direct transposition by both targeted and untargeted mechanisms, I can use comparisons between these mechanisms that distinguish between factors that affect transposition in general and transposon targeting alone. Other aspects of Tn7 transposition also provide multiple means of analysis. Among the observable effects are transposition frequency, transposon target profile, SOS induction (affected by either TnsE alone or excision of the Tn7 element), genomic island formation (or loss of target-site immunity), and strictly biochemical effects such as

excision and end joining can be observed in vitro. In this Chapter I will discuss the analysis of the frequency of TnsE-mediated transposition in selected mutant backgrounds to better understand the basis of transposition into the chromosome. The experiments in this chapter are limited to transposition frequency, but with an eye toward future experiments that may involve combinations of the above mentioned types of analyses of Tn7 transposition pathways. In these experiments I compare the effects of specific mutations on the frequency of TnsABC+E transposition with the mutant TnsABC* transposition pathway which contains a class I mutation in TnsC ($TnsC^{A225V}$) (see Chapter 1) (Stellwagen and Craig, 1997; Stellwagen and Craig, 1998). TnsABC* selects transposition targets at random and serves as a measure of the affect of the cellular milieu on the transposon core machinery alone, providing an important positive control and an indicator of cell viability. Comparison between the two pathways allows analysis of affects that are specific to TnsE and not transposition in general. In this chapter, I will address the effects of homologous recombination, DNA damage response, replication restart, and DNA methylation on TnsABC+E transposition.

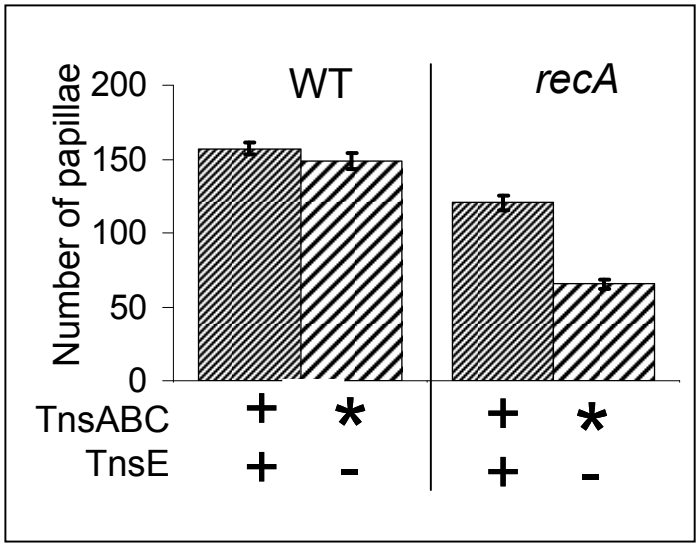
4.3 Results

Homologous recombination is not required for TnsE-mediated transposition.

Since DNA double-strand breaks are a preferred target for TnsABC+E transposition, it would make sense if the target that is identified by TnsE is specific to homologous recombination. In *E. coli* double-stranded DNA breaks (DSBs) are either fixed by homologous recombination or degraded by RecBCD. Many experiments involving TnsABC+E transposition have been

done in *recA*(-) cells, so it has been clear for some time that there is no requirement for homologous recombination in TnsE-mediated transposition. But, what affect does *recA* have on transposition in the TnsE pathway? I used strains containing a miniTn7 element in the *attTn7* locus in a papillation assay (see Experimental procedures) to determine the transposition frequency in a *recA* background. I found that the transposition frequency in general is decreased (Figure 4.1). The effect was not specific to TnsE-mediated transposition because TnsABC* transposition also showed a comparable, if not greater, reduction in transposition frequency. The TnsABC* core machinery has a mutant $TnsC^{A225V}$ that does not require TnsE for targeting or activation (Stellwagen and Craig, 1997). This result suggests that the lack of homologous recombination does not specifically hinder TnsE-mediated transposition. The decrease in transposition frequency in both TnsABC+E and TnsABC* pathways is likely due to the inability to repair DSBs left by excision of the element from the *attTn7* site rather than deficiencies in targeting the transposon. This deficiency may reduce background because transposition events that occur into the same sister chromosome that sustains the DSB from the excision of the transposon would be expected to be degraded by the RecBCD enzyme, or otherwise lost, making those events silent in this assay. It is interesting to note that TnsABC* showed a greater reduction in the transposition frequency. This may indicate that the deregulation of transposition results in behavior of the transposon that is a greater detriment to the host cell or to the transposon than a comparable level of targeted transposition. The observation that random transposition frequency is reduced to a greater level than targeted transposition frequency is consistent with the view that TnsE-mediated transposition occurs in way that

Figure 4.1. TnsABC+E transposition in *recA*(+) and *recA*(-) genetic backgrounds. In the *recA* background transposition is impaired for both TnsE-mediated and, to a greater extent, TnsABC* transposition, indicating that the reduction in transposition is not specific to TnsE. Reduction in transposition frequency may be due to trouble repairing DSBs left in the *attTn7* locus by excision of the mTn7 element. The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).


Number of papillae
200
150
100
50
0
WT
recA
TnsABC
TnsE
+
*
+
-
+
*
+
-

protects the integrity of the genome in ways that random transposition does not (see Chapter 1).

Constitutive activation of the LexA regulon does not stimulate TnsE-mediated transposition.

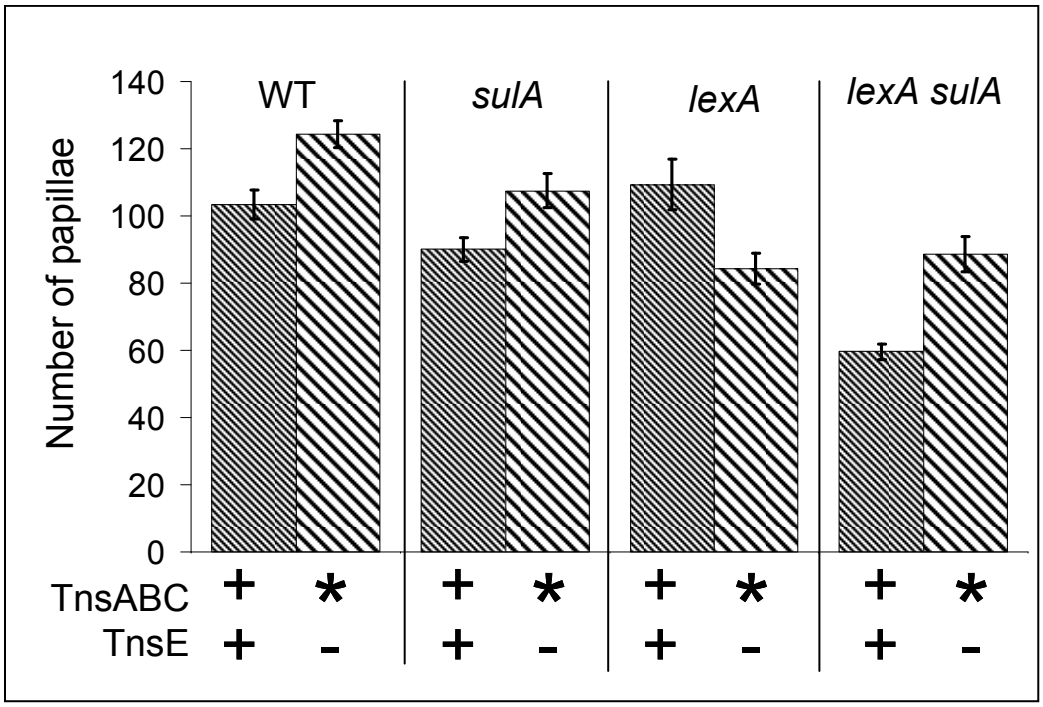
There are many types of DNA damage that stimulate TnsE-mediated transposition that also stimulate the SOS response (Shi, 2008). It is therefore possible that induction of the SOS response might actually be responsible for the increase in TnsE-mediated transposition. To test this idea I monitored transposition in a strain background that did not have the LexA protein that is responsible for regulating genes under the SOS response (Experimental procedures). To accurately gauge the effect of the SOS response I also included a *sulA* allele; induction of SulA during the SOS response results in the formation of cell filaments, which may interfere with the papillation assay. I found that constitutive SOS induction in the *lexA* backgrounds does not stimulate TnsE-mediated transposition (Figure 4.2). There is a significant drop ($p<0.0005$, T-test, two tailed, equal varience) in all transposition in the *sulA lexA* background, which may be explained by any of a number of genes in the SOS regulon effecting cell viability. Constitutive activation of genes in the SOS regulon may pre-dispose cells to dealing with endogenous DNA damage rapidly, perhaps limiting the amount of time that TnsE has to recognize a structure or complex that is associated with DNA damage repair.

TnsE-mediated transposition does not specifically require any one of the known replication restart pathways.

TnsE is specifically stimulated by the repair of DSBs and the repair of DNA damage caused by UV irradiation, mitomycin C, and phleomycin, all of which require a mechanism to restart DNA replication forks (Shi, 2008). Given the requirement for replication restart in the repair of double-stranded DNA breaks it appeared possible that TnsE required one of the proteins in the replication restart pathway to recognize target DNAs. The attraction of TnsE-mediated transposition to the terminus region could also be explained by an interaction between this pathway and the replication restart apparatus (Bidnenko et al., 2006). Replication fork restart can occur through a number of genetically distinct pathways that likely help address the various ways a DNA replication fork can stall or collapse (Heller and Marians, 2006; Sandler, 2000). I determined if any of the multiple pathways for replication fork restart was specifically required for TnsE-mediated transposition during normal growth.

PriA is a 3' to 5' helicase that is required for the major replication fork restart pathways in *E. coli* (Sandler and Marians, 2000). I monitored transposition in a *ΔpriA* strain and found that this component of the restart apparatus was not required for TnsE-mediated transposition (Figure 4.3). Surprisingly, I found that TnsE-mediated transposition was actually stimulated in *ΔpriA* strains. This effect was specific to TnsE-mediated transposition because no change was found in TnsABC* transposition; the TnsABC* core machinery has a mutant TnsC^{A225V} that does not require TnsE for targeting or activation (Stellwagen and Craig, 1997). It is possible that the stimulatory effect found in the *ΔpriA* background might be caused by TnsE favoring a

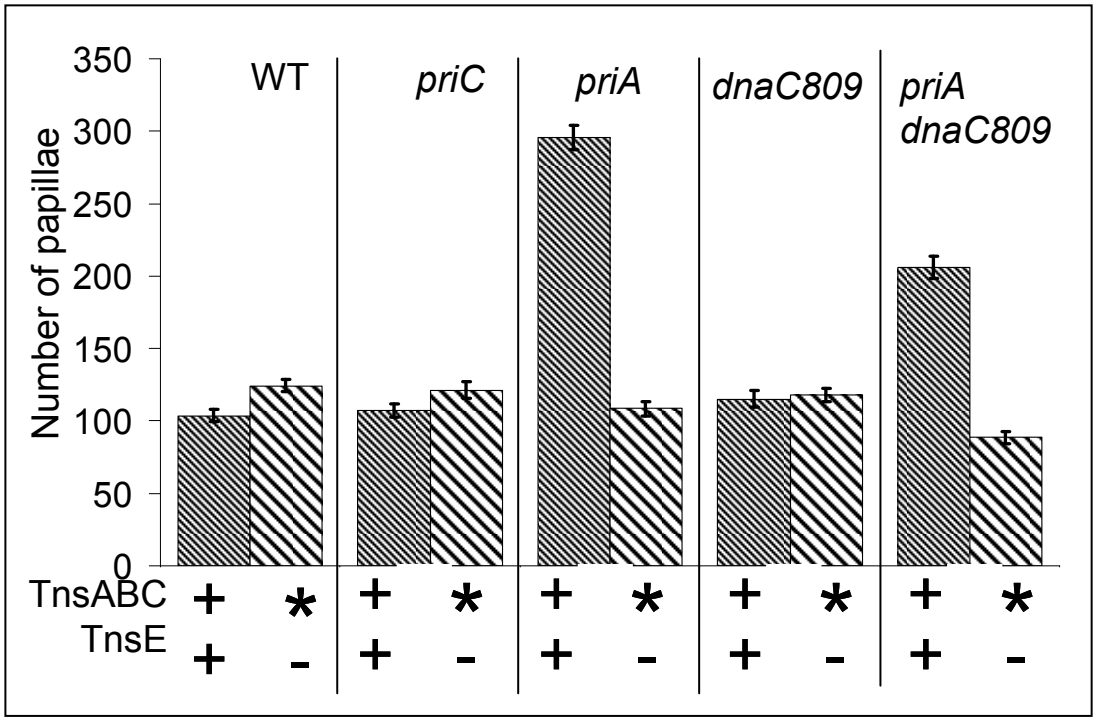
Figure 4.2. TnsABC+E transposition in *lexA, sulA,* and *lexA*(-) genetic backgrounds. Comparison of transposition in *lexA* and *lexA*(-) backgrounds indicates that constitutive expression of SOS induced genes does not stimulate TnsE-mediated transposition. Deletion of the *sulA* gene prevents filamentation of *E. coli* cells that lack *lexA,* ruling out affects associated with filamentation that may confound the papillation assay. TnsABC* is included as a positive control to rule out effects that are not associated with TnsE targeted transposition (*). The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).


Number of papillae
140
120
100
80
60
40
20
0
WT
sulA
lexA
lexA sulA
TnsABC
TnsE
+
*
+
-

structure or complex that is processed more slowly in a *ΔpriA* background. To further rule out any requirement for PriA, I utilized a gain of function mutation in *dnaC*, called *dnaC809,* which allows replication restart in a *ΔpriA* background. DnaC loads the replicative helicase, DnaB, as part of initiating all DNA replication forks (Kornberg and Baker, 1992). The *dnaC809* allele bypasses the need for multiple other replication fork restart proteins like PriB and DnaT, *in vivo* and *in vitro* (Liu et al., 1999; Sandler et al., 1996). TnsABC+E transposition was unaffected by the *dnaC809* allele alone (Figure 4.3). The results found with the *ΔpriA* and *dnaC809* strains indicates that the replication fork restart pathway that requires the PriA, PriB and DnaT proteins is not required for TnsE transposition (i.e. a specific interaction between TnsE and one of these proteins or a specific structure made by these proteins is not required for transposition *in vivo*). I found that the stimulation of TnsABC+E transposition found in the *ΔpriA* strain was not suppressed by the *dnaC809* allele (Figure 4.3). The inability of *dnaC809* to suppress the stimulation in transposition is consistent with the observation that this allele of *dnaC* is not as efficient at restarting replication forks as the *priA* pathway (Sandler et al., 1999).

PriA is essential for the major replication fork restart pathways involving PriB and a second pathway involving PriC (Sandler, 2005). However, PriC can also restart DNA replication forks in a *ΔpriA* background through the use of a second DNA 3' to 5' DNA helicase, called Rep. The *dnaC809* suppressor also requires *priC* and *rep* to suppress mutations in *priA* (McCool and Sandler, 2001). I monitored transposition in a *ΔpriC* strain to determine if the PriC-Rep pathway of replication fork restart was required for TnsE-mediated transposition (Figure 4.3). I found that TnsABC+E transposition is unaffected

Figure 4.3. TnsABC+E transposition in *priA*, *priC*, and *dnaC809* genetic backgrounds. The PriA and PriC proteins are not required for TnsABC+E transposition, but is sensitive to which replication restart pathway is used. The *dnaC809* allele produces a mutant helicase loading enzyme that does not require PriA or PriC. The TnsABC* mutant core machinery that allows untargeted transposition is included as a positive control to rule out effects that are not associated with TnsE targeted transposition (*). The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).

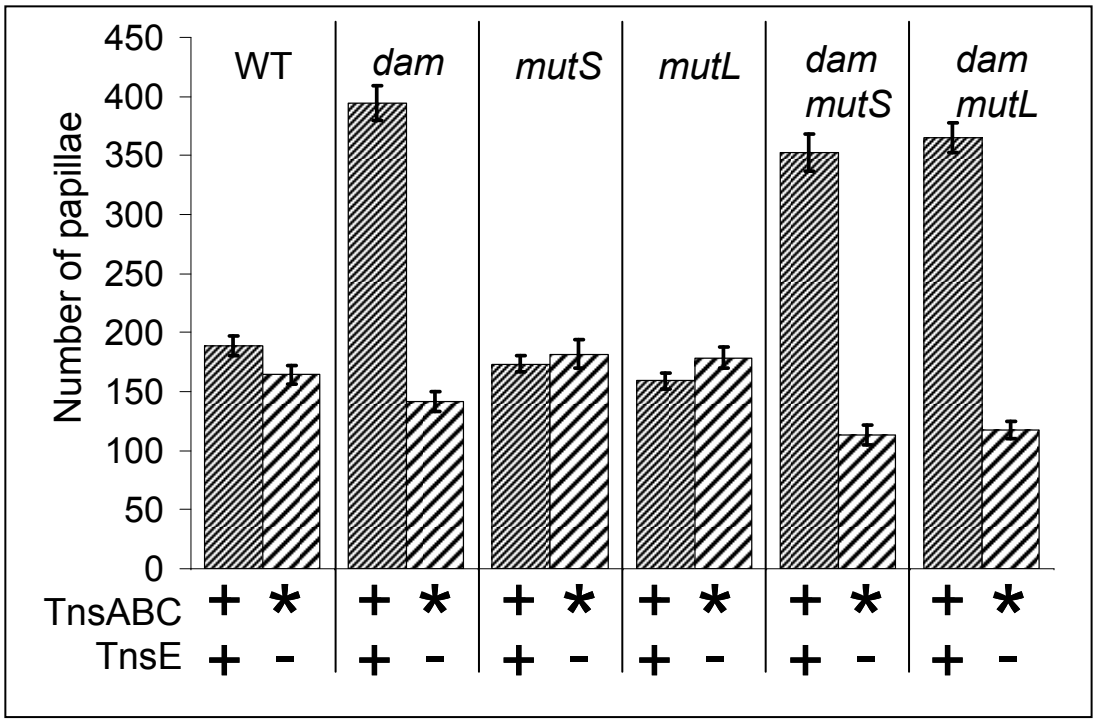

Number of papillae
350
300
250
200
150
100
50
0
WT
priC
priA
dnaC809
priA dnaC809
TnsABC
TnsE
+
*
+
-

by the *priC* allele when compared with the TnsABC* control strain. Taken together, the observation that PriA and PriC are not required for transposition indicates that no protein that is required for these pathways is essential for TnsE-mediated transposition. This includes the known pathways that require PriB, DnaT, and Rep (McCool and Sandler, 2001). It is unknown why TnsABC+E transposition is stimulated in *ΔpriA* strains, but this could indicate that slowed DNA repair allows a structure or complex that is preferred by TnsE to linger in the cell.

Loss of Dam methylation stimulates TnsE-mediated transposition by an unknown mechanism.

Early genetic screens for mutations that affected TnsE-mediated transposition revealed that knocking out the Dam methylase stimulates TnsE-mediated transposition (DeBoy, 1997). Under the conditions of the papillation assay in which TnsABC+E are produced from a *lac* promoter, transposition is stimulated two- to three-fold in a *dam* background (Figure 4.4, 4.5). The effect is specific to TnsABC+E transposition; there is no stimulation in TnsABC* transposition in this background. Dam adds a methyl group to the N-6 position of adenines in GATC nucleotide sequences. The *dam*(-) mutation displays a diverse range of phenotypes due to the reliance of many systems on the methylation of DNA. Cells that lack the *dam* gene are chronically induced for the SOS response, accumulate DNA double-stranded breaks, and require the RecA protein for survival. Some organisms, such as *Vibrio cholerae* and *Yersinia pestis,* absolutely require Dam for survival (Lobner-Olesen et al., 2005). Dam methylation affects gene transcription, methyl-directed mismatch

Figure 4.4. TnsABC+E transposition in *dam* and MMR deficient backgrounds. TnsE-mediated transposition is specifically stimulated in the *dam*(-) genetic background (i.e. there is no corresponding stimulation of the TnsABC* non-targeted transposition pathway). This stimulation is not due to an increased frequency of DSBs. Strains lacking the Dam protein accumulate DSBs in a mismatch repair dependent fashion. Elimination of MutL or MutS does not suppress the *dam*(-) associated TnsABC+E stimulation. The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).


Number of papillae
450
400
350
300
250
200
150
100
50
0
WT
dam
mutS
mutL
dam mutS
dam mutL
TnsABC + * + * + * + * + * + *
TnsE + - + - + - + - + - + -

repair (MMR), the synchronization of replication initiation, and the structuring of the nucleoid following DNA replication (Lobner-Olesen et al., 2005). In addition to these effects within a host cell genome, DNA methylation by Dam also affects the activity of some transposons, such as Tn10 and Tn5, by modulating transposase expression and the activity of transposon ends (Kleckner, 1990; Reznikoff, 1993). Previous experiments focused on the direct affect of Dam methylation on Tn7 transposition by mutation of each of the GATC sites in the transposon ends. These experiments did not show any difference in transposition (DeBoy, 1997).

DNA double-strand breaks accumulate in *dam*(-) cells for a number of reasons. The first reason is that DSBs accumulate in *dam* cells is due to the deregulation of the MMR system. In *E. coli*, MutS detects mismatches in DNA sequences, recruits MutL and MutH, and activates the nicking activity of MutH (Kunkel and Erie, 2005). MutH is sensitive methylation of DNA at GATC site and only nicks the more recently replicated (non-methylated) strand of DNA. When DNA is persistently non-methylated, MutH will nick non-replicated strands leading to single-stranded gaps which are eventually converted to double-strand breaks by a second replication fork if not fixed promptly (Nowosielska et al., 2006). Double-strand breaks (DSBs) may also be the direct result of MutH nicking both strands of DNA (Au et al., 1992). This is illustrated by the fact that rapidly growing cells accumulate more DSBs than do slowly growing cells in a *dam recB*^{ts} background (Nowosielska and Marinus, 2008). Since TnsE has been shown to specifically target DSBs for insertion, I hypothesized that the stimulation observed in *dam*(-) cells was due to accumulation of double-strand breaks. Cells that are *dam*(-) and deficient in

MMR do not accumulate DSBs to the level of *dam* alone (Robbins-Manke et al., 2005). I constructed *dam mutS* and *dam mutL* double mutants to test the hypothesis that DSBs cause the stimulation in TnsE-mediated transposition. The formation of DSBs by MutH is dependent on the present of both MutS and MutL (Nowosielska and Marinus, 2008). I expected to see a decrease in stimulation of transposition in the double mutants due to the reduction in DSB formation. I did not see a significant decrease in TnsE-mediated transposition, suggesting that some other factor associated with the lack of Dam methylation was the cause of the stimulation (Figure 4.5). This results was surprising, since elimination of MMR by deleting of *mutS* drastically reduces dam cell's level of endogenous DNA double-stranded breaks, from an average of 1.2 per cell to an average of 0.4 per cell (as measured by microgel electrophoresis) (Robbins-Manke et al., 2005). The level of DSBs per cell is still higher than that found in wild-type (~0.2 breaks per cell) but still within the error range as compared to *mutS* alone (~0.3 breaks per cell). It is not likely that the stimulation in TnsE-mediated transposition in the *dam*- background is a result of DSB formation, since there is no difference in TnsABC+E transposition frequency when comparing *dam mutS* cells and *dam* alone cells, but far more DSBs in *dam* alone as compared to *dam mutS* strains.

The influence of Dam methylation on the SeqA protein might also provide and explanation for the increased TnsABC+E transposition frequency. The SeqA protein is responsible for "sequestering" newly replicated DNA to prevent over-initiation of the origin (Slater et al., 1995). SeqA also helps organize the nucleoid following replication and can be seen associated with the replication fork in cytological studies (Fossum et al., 2007).

SeqA binds preferentially to hemimethylated DNA, and prevents rapid methylation (Slater et al., 1995). Null mutants of *seqA* and *dam* are epistatic in some DNA damage repair assays (Sutera and Lovett, 2006). Cells that lack SeqA proteins display asynchronous timing of replication initiation (Camara et al., 2005) and it has been proposed that this is the main reason for the sensitivity to DNA damaging agents in *seqA* strains (Sutera and Lovett, 2006). I initially tried to directly test the affect of *seqA* null mutations on TnsE-mediated transposition, but found that these strains are incompatible with both the papillation assay and the lambda hop assay. No papillae can be observed even in positive TnsABC* controls (data not shown). The lambda hop assay was not used because of the requirement of the assay to use *recA*(-) cells; *seqA* is synthetically lethal with *recA* (Kouzminova et al., 2004). Since the loss of SeqA function in *dam*(-) cells leads to asynchronous origin firing, I decided to directly test if asynchronous DNA replication caused stimulation in TnsE-mediated transposition. Asynchronous initiation of replication can lead to the formation of double-stranded DNA breaks due to head-to-tail replication fork collisions that result in replication fork collapse (Sutera and Lovett, 2006). This is thought to be the main reason that *seqA* strains are sensitive to DNA damaging agents and require the RecA protein. I used the temperature sensitive *dnaA46* allele, which causes a very high degree of asynchronous origin firing, and found that there was no detectable increase in TnsE-mediated transposition (Figur 4.5) (Katayama et al., 1997). While improper functioning of the SeqA in *dam*(-) cells cannot be ruled out as the cause of increased TnsE-mediated transposition frequency, this experiment has ruled out a specific effect of asynchronous regulation of replication initiation as the causative agent. It is possible that other functions of SeqA, such as nucleoid

organization, affect transposition specifically in the TnsE pathway.

The *dnaA46* allele makes cells very sick, making papillation assay results difficult to interpret, so I decided to test another mutation at the same time that causes an asynchrony phenotype. Hda is a protein that helps regulate the cell cycle by deactivating the DnaA protein through stimulation of its ATPase activity. This process is referred to as regulatory inactivation of DnaA, or RIDA. Hda must interact with the processivity factor that remains on double-stranded DNA to achieve DnaA inactivation. Interestingly, I found that the TnsE-mediated transposition frequency was stimulated in the *hda* background, nearly two fold ($p<0.0005$, T-test, equal variance, two-tailed) (Figure 4.6). When compared side by side with *dam* strains, *hda* does not stimulate TnsABC+E transposition as highly as *dam* (Figure 4.6). Hda and TnsE both bind to the β clamp. It is possible that by eliminating the Hda protein there is less competition for the clamp, allowing TnsE to bind to the clamp more frequently and target transposition. In this case, it might be expected that more transposition events occur near the origin, and possibly near the *dat* locus (a region of the chromosome that contains many *dnaA* binding sites) (Camara et al., 2005; Kaguni, 2006), since these are the regions of the chromosome where RIDA is expected to be focused. It is also conceivable that *hda* binds to clamps throughout the cell cycle at every site on the chromosome, and participates in an ongoing competition for access to the processivity factor. In this case, transposition insertions would not necessarily be expected to concentrate at the origin or at the *dat* locus. There is no transposon mapping data for *hda* or *dam* strains, so it is not know where in the chromosome these transposition events are occurring. Mapping transposon insertions in both of these backgrounds will be essential for gaining a more

Figure 4.5. Comparison of TnsABC+E transposition in the *dam* and *dnaA46* genetic backgrounds. Asynchronous (over-initiated) DNA replication in a *dam*(-) background is not the cause of the stimulation in TnsE-mediated transposition. A hypomorphic *dnaA46* allele is known to cause over-initiation of replication, yet does not stimulate transposition. The TnsABC* mutant core machinery that allows untargeted transposition is included as a positive control to rule out effects that are not associated with TnsE targeted transposition (*). The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).

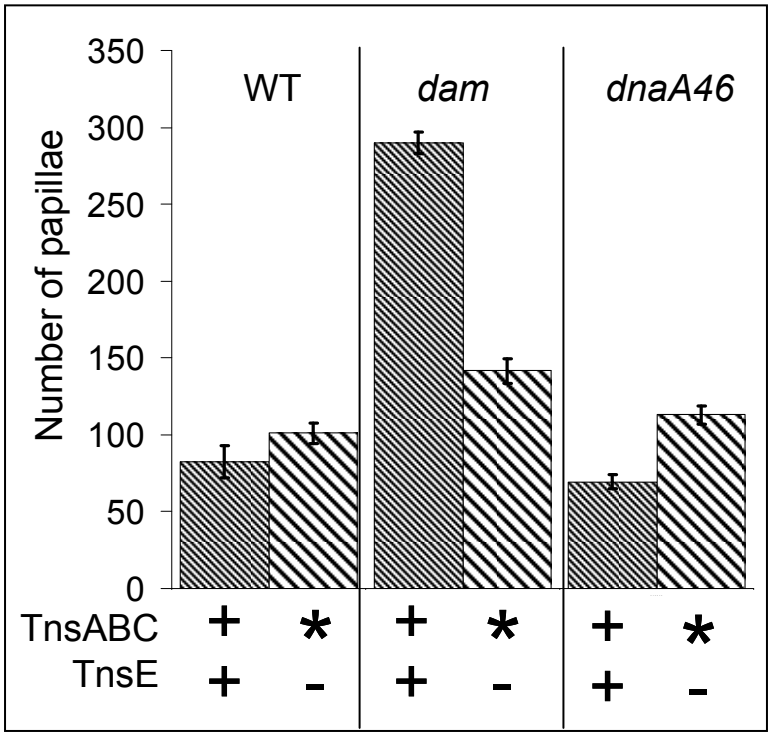
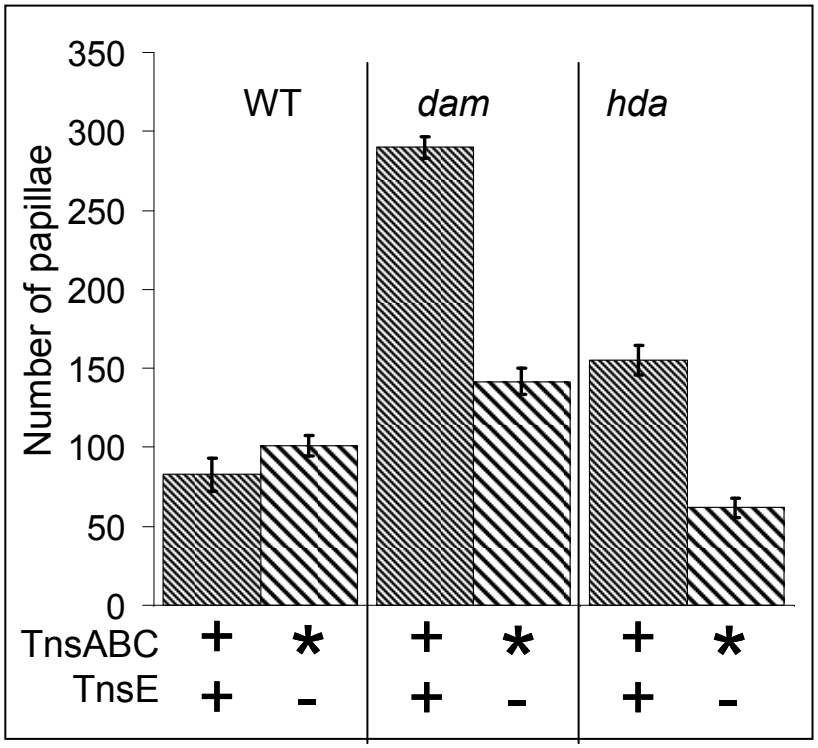

350
300
250
200
150
100
50
0
Number of papillae
WT
dam
dnaA46
TnsABC
TnsE
+
*
+
*
+
*
+
-
+
-
+
-

Figure 4.6. Comparison of TnsABC+E transposition in *dam* and *hda* genetic backgrounds. Deletion of the *hda* gene also result in a mild stimulation of TnsABC+E transposition, possibly due to altered β clamp activity. The stimulation in TnsABC+E transposition frequency is not do to the asynchronous phenotype observed in hda strains (see text and Figure 4.5). The TnsABC* mutant core machinery that allows untargeted transposition is included as a positive control to rule out effects that are not associated with TnsE targeted transposition (*). The frequency of transposition is indicated on the Y-axis as the number of Lac^+ papillae after 72 hours at 30˚C on MacConkey's media. Error bars indicate the standard error of the mean (n=12).


Number of papillae
350
300
250
200
150
100
50
0
WT
dam
hda
TnsABC
TnsE
+
*
+
-

complete picture of the probable cause of TnsABC+E transposition stimulation.

4.4 Discussion

In this Chapter, I have shown that TnsE-mediated transposition does not specifically rely on the homologous recombination process, or replication restart to efficiently stimulate transposition in the absence of a conjugal plasmid. In fact, I find that elimination of the *priA* gene specifically stimulates transposition, but *priC* does not indicating that TnsE is sensitive to which replication restart pathway is utilized. Repair pathways involving PriA or PriC must restart DNA replication when forks stall or collapse at DNA damage (Heller and Marians, 2005). Strains with the *ΔpriA* allele must use the PriC pathway, which may be less efficient at restarting replication on certain kinds of substrates, possibly resulting in persistently stalled replication forks somewhere in the chromosome (Heller and Marians, 2005). The preferred PriA substrates are replication forks where the lagging-strand has stalled (Heller and Marians, 2006). PriA is also capable of restarting a replication ahead of a stalled replication fork, preventing a stalled fork from holding back the replication of the rest of the chromosome. In the absence of PriA, a stalled fork may persist for a longer time, increasing the probability that TnsE will be able to bind to clamps remaining on DNA from the stalled fork (Heller and Marians, 2005). The finding that TnsE-mediated transposition is stimulated by the ΔpriA mutation but not the ΔpriC mutation may indicate that TnsE may be more sensitive to lesions on the lagging-strand than on the leading-strand. Perhaps the decoupling of leading- and lagging-strand DNA synthesis observed in ΔpriA mutants (Heller and Marians, 2005), and stalled replication

machinery on the lagging-strand exposes a complex, such as a β clamp juxtaposed with a single-strand DNA interruption (see chapter 3, Figure 3.17), that may be specifically recognized by TnsE.

Stimulation of TnsE-mediated transposition in a *dam*(-) background remains mysterious. While inducing DSBs at specific locations has been shown to stimulate TnsE-mediated transposition, the formation of MMR induced DSBs in *dam* strains is not the cause of increased transposition frequency. I also ruled out another model that might explain the stimulation in TnsABC+E transposition by a mechanism involving increased DSBs or other affects due to asynchronous initiation of DNA replication. These observations hint at a characteristic or defect of *dam*(-) cells that may not be clear at this time. TnsE may be capable of identifying a structure or complex in *dam* cells that otherwise remains obscure. Given the experiments mentioned in Chapter 3, factors that are expected to increase the frequency of TnsE-mediated transposition are thought to increase available β clamps by stalling replication forks or repair associated DNA replication. Alternatively, TnsE may be stimulated by the product of a gene whose regulation is sensitive to the methylation state of its promoter (Lobner-Olesen et al., 2005).

In an experiment designed to test the hypothesis that asynchronous DNA replication might cause an increase in TnsABC+E transposition frequency, I tested to see if deletion of the *hda* gene also stimulated TnsE-mediated transposition. There was a mild stimulation in TnsE-mediated transposition and a decrease in TnsC*-mediated transposition. In conjunction with the result observed with the *dnaA46* allele, this suggests that there is a stimulatory effect associated with the loss of Hda that is not associated with asynchrony. Hda binds to the β clamp, preferentially when it remains on

double-stranded DNA, as opposed to multiply primed DNA substrates (Su'etsugu et al., 2004). This substrate may be similar to a substrate on which TnsE is predicted to initially interact with β (see Chapter 3, Figure 3.17). It is possible that the β clamp is more available to bind to TnsE when *hda* is not present to compete with TnsE on the same DNA substrate.

Some mutations described above display a greater decrease in TnsABC* transposition frequency compared to TnsABC+E (Figure 4.1, 4.2, 4.6). These results may provide evidence that regulation of transposition protects the host cell and the transposon in some way that is not present in the case of random transposition. All of the assays presented above where done in cells in which the Tn7 donor is within the bacterial chromosome itself (the miniTn7 element was situated within the attTn7 locus). Since TnsE senses ongoing replication and functioning replication repair, TnsE may ensure that DSBs in the chromosome that are created by transposition are able to be repaired.

Taken together with the data presented in Chapter 3, the results above may indicate that deletion of *hda*, *priA*, and possibly *dam* all result in a condition in which β clamps remain unbound by other proteins on double-stranded DNA, yet proximal to a single-stranded gap. Experiments conducted by Georgescu et al. suggest that β may physically interact with the first four bases of single-stranded DNA at the the 3' end at single-stranded interruptions, holding the clamp in place and recruiting repair proteins to the site of the gap (Georgescu et al., 2008). Others have shown that the β clamp may interact with single-strand binding protein (SSB) to remain at the primer template junction (Laurence et al., 2008). The presence of a forked structure instead of a gap may prevent β from interacting at the 3' end, at which point it

may slide along DNA and become a substrate for TnsE binding. TnsE may be more able than β to bind to 3' ends at forked structures. Indeed, it has been shown that TnsE can bind to forked structures with gaps as small as 5 bp, the smallest gap tested in this system (Peters and Craig, 2001).

Since TnsE-mediated transposition is specifically stimulated by, and preferentially directs transposition into, the discontinuous DNA replication process that occurs as conjugal plasmids enter a new host cell it is reasonable to presume that TnsE-specific stimulation occurs when leading- and lagging strand DNA replication are uncoupled. Since TnsE uses its interaction with β to direct transposition to these targets, it seems likely that β is especially exposed under these conditions. TnsABC+E transposition frequency may serve as an indicator of the frequency of uncoupling between leading- and lagging-strand DNA synthesis. One might be able to determine where replication terminates in these cases by mapping the transposon insertions. The orientation of Tn7 ends provides information about the direction of repair associated replication. It will be interesting to map transposon insertions in the mutant backgrounds described above and compare them to insertions found in wild-type chromosomes to determine if there are any differences in orientation or target-sites that help explain the increases in transposition frequency. This type of analysis would also be useful in understanding how DNA damage associated replication is coordinated.

4.5 Experimental procedures

Bacterial strains and plasmids

Strains were constructed using standard P1 transduction methods (Miller, 1992; Peters, 2007) (Table 4.1). Drug markers were removed from *frt*-

containing constructs with the FLP recombinase produced from pCP20, which was subsequently lost at 42° (Datsenko and Wanner, 2000). Transduction of the *dnaC809* allele was by linkage to *thrA-34*::Tn10 and confirmed by PCR amplification folloId by restriction digestion with *Hin*FI (the *dnaC809* mutation abolishes a *Hin*FI site within the *dnaC* gene)(SANDLER ET AL., 1996). Strain AP457 was constructed by the method described by Datsenko and Wanner (Datsenko and Wanner, 2000). Briefly, primers JEP191 (5'-AGA AAA ACT CAT CGA GCA TCA AAT GAA ATG AAA CTG CAA TTT ATT CAT A GT GTA GGC TGG AGC TGC TTC-3') and JEP192 (5'-ATG AGC CAT ATT CAA CGG GAA ACG TCT TGC TCG AGG CCG CGA TTA AAT TCA TAT GAA TAT CCT CCT TAG-3') Were used to amplify the Cam^R gene (and flanking *frt* sites) from plasmid pKD3 with 40 bp homology to Kan^R gene cassettes. The PCR product was crossed into AP429 carrying pKD46, which was induced with 0.2% arabinose and the plasmid subsequently lost by growth at 42° giving AP453. Plasmid vectors were constructed with standard techniques as described (Sambrook et al., 1989)(Table 4.2).

Transposition assay

A promoter capture assay, called the papillation assay, monitors transposition levels in lawns of bacteria on indicator media (Huisman and Kleckner, 1987; Stellwagen and Craig, 1997). A miniTn7 element encoding the lactose utilization genes without the requisite promoter is situated in the tester strain *attTn7* site (JP617, AP429, or AP457) (Table 1). In this configuration the *lac* genes are not transcribed and the strain is unable to use

TABLE 4.1 Strains used in this study

Strain	Genotype	Reference
MC4100	*araD139 Δ(argF-lac)169 rpsL150 relA1 flhD5301 deoC1 ptsF25 rbsR22 e14⁻ Δ(fimB-fimE)632::IS1 Δ(fruK-yeiR)725*	(Casadaban, 1976; Peters et al., 2003)
NLC28	MC4100 Val^R	(McKown et al.,
JP617	NLC28 *attTn7*::miniTn7(R90-*lacZYA'*-Kan^R-L166)	(McKown et al., 1988; Peters and
JP1588	JP617 *priA*::Cam^R	Nancy Craig
AP44	JP617 *dnaC809*	P1(JC19008) X
AP48	JP617 *dnaC809 priA*::Cam^R	P1(JP1588) X AP44
JP1386	NLC28 *Δara714*	(Peters and Craig,
AP429	JP1386 *attTn7*::miniTn7(R90-*lacZYA'*-Kan^R-L166)	P1(JP617) X JP1386
AP453	JP1386 *attTn7*::miniTn7(R90-*lacZYA'-frt*-Cam^R-*frt*-L166)	This work
AP457	JP1386 *attTn7*::miniTn7(R90-*lacZYA'-frt*-L166)	This work
AP523	AP457 *priC303*::Kan^R	P1(JC19008) X AP457
AP767	AP457 *ΔsulA6209::*Tet^R	P1(GM7542) X AP457
AP964	AP457 *lexA71::Tn5*	P1(KL788) X AP457
AP965	AP457 *lexA71::Tn5 ΔsulA6209::*Tet^R	P1(KL788) X AP767

TABLE 4.1 (Continued)

Strain	Genotype	Reference
AP40	JP617 *thrA-34*::Tn10	P1(CAG18442) X
AP223	JP617 *dam13::*Tn9	P1(ER2925) X JP617
AP226	JP617 *mutS::*Tn10	P1(GM2165) X JP617
AP229	JP617 *mutL::*Tn10	P1(GM2166) X JP617
AP235	JP617 *mutS::*Tn10 *dam13::*Tn9	P1(ER2925) X AP226
AP238	JP617 *mutL::*Tn10 *dam13::*Tn9	P1(ER2925) X AP229
GM2165	F^- *dcm-6 thr-1 ara-14 leuB6 tonA31 lacY1 tsx-78 supE44 galK2 galT22 hisG4 rpsL136*	Martin Marinus
GM2166	F^- *dcm-6 thr-1 ara-14 leuB6 tonA31 lacY1 tsx-78 supE44 galK2 galT22 hisG4 rpsL136*	Martin Marinus
JC326	MG1655 *seqA*::TetR	Elliott Crooke (Camara et al., 2005)
JC125	MG1655 *hda*::TetR	Elliott Crooke (Camara et al., 2005)
KA452	*thyA thr tyrA(Am) trpE9829(Am) metE deo dnaA46ts tnaA::*Tn10	Elliott Crooke (Camara et al., 2005)
AP334	JP617 *seqA::*TetR	P1(JC326) X JP617
AP332	JP617 *hda::TetR*	P1(JC125) X JP617
AP340	JP617 *dnaA46 tnaA*::Tn10	P1(KA452) X JP617

TABLE 4.1 (Continued)

Strain	Genotype	Reference
BW25113	MG1655 *Δ(araD-araB)567 lacZ-4787Δ(::rrnB-3) Δ(rhaD-rhaB)568 hsdR514*	(Datsenko and Wanner, 2000)
GM7542	MG1655 *ΔsulA6209::*Tet^R	Martin Marinus via Mark Sutton
JC19008	JC18983 *dnaC809 thrA*$^+$ *priC303*::Kan^R	(Sandler et al., 1996)
KL788	*Δ(gpt-lac)5 tsx-35 sulA3 e14- rfbC1 recA441(ts) relA1 rpsL31 kdgK51 mtl-1 spoT1 thi-1 lexA71::Tn5 creC510*	(Miller et al., 1981)
ER2925	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::*Tn10*)TetS endA1 rpsL136 dam13::*Tn9 *xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New England Biolabs

TABLE 4.2 Plasmids used in this study

Name	Relevant information
pTA106	pSC101 replicon, ampicillin resistant cloning vector (Peters and Craig, 2000)
pQS100	pTA106 vector encoding TnsABC (Shi, 2008)
pQS102	pTA106 vector encoding TnsABC+E (Shi, 2008)
pQS107	pTA106 vector encoding $TnsABC^{A225V}$ (Shi, 2008)
pCP20	Temperature sensitive plasmid with thermal induction of FLP recombinase (Cherepanov and Wackernagel, 1995)
pKD46	Temperature sensitive plasmid with the λ Red proteins under arabinose control (Datsenko and Wanner, 2000)
pKD3	Plasmid encoding ampicillin resistance that allows PCR amplification of a gene cassette encoding chloramphenicol resistance flanked by *frt* sites recognized by the FLP recombinase (Datsenko and Wanner, 2000)

lactose (Lac^-). If the element transposes to a new position in the correct orientation in an active gene the element will be transcribed resulting in a phenotypically Lac^+ bacterium. On MacConkey's lactose indicator media or on LB lactose X-gal media, Lac^+ cells form differentially colored microcolonies, or papillae, on the lawn of Lac^- cells. The number of Lac^+ papillae provides a measure of transposition levels. To rule out any effect of the tested allele or condition on cell viability or papillae formation an internal control is always tested where papillae formation is compared with a background strain that lacks TnsE.

4.6 Acknowledgements

I thank Nancy Craig (Johns Hopkins Medical School, Howard Hughes Medical Institute), Susan Lovett (Brandeis University), Steve Sandler (University of Massachusetts at Amherst), Mark Sutton (University at Buffalo), and the *E. coli* Genetic Stock Center, Yale University for providing strains and plasmids. I thank Zaoping Li and Qiaojuan Shi for comments. This work was funded by a grant from the National Science Foundation (MCB-0315316).

REFERENCES

Au, K. G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. J Biol Chem *267*, 12142-12148.

Bidnenko, V., Lestini, R., and Michel, B. (2006). The *Escherichia coli* UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. Mol Microbiol *62*, 382-396.

Camara, J. E., Breier, A. M., Brendler, T., Austin, S., Cozzarelli, N. R., and Crooke, E. (2005). Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of *Escherichia coli* DNA replication. EMBO Rep *6*, 736-741.

Casadaban, M. J. (1976). Transposition and fusion of the lac genes to selected promoters in Escherichia coli using bacteriophage lambda and Mu. J Mol Biol *104*, 541-555.

Cherepanov, P. P., and Wackernagel, W. (1995). Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9-14.

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

DeBoy, R. T. (1997) Transposon Tn7: target DNAs can modulate Tn7 insertion, Ph.D. Thesis., Johns Hopkins University., Baltimore, Maryland.

Fossum, S., Crooke, E., and Skarstad, K. (2007). Organization of sister origins and replisomes during multifork DNA replication in *Escherichia coli.* Embo J *26*, 4514-4522.

Georgescu, R. E., Kim, S. S., Yurieva, O., Kuriyan, J., Kong, X. P., and O'Donnell, M. (2008). Structure of a sliding clamp on DNA. Cell *132*, 43-54.

Heller, R. C., and Marians, K. J. (2005). The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. Mol Cell *17*, 733-743.

Heller, R. C., and Marians, K. J. (2006). Replisome assembly and the direct restart of stalled replication forks. Nat Rev Mol Cell Biol *7*, 932-943.

Huisman, O., and Kleckner, N. (1987). A new generalizable test for detection of mutations affecting Tn10 transposition. Genetics *116*, 185-189.

Kaguni, J. (2006). DnaA: Controlling the Initiation of Bacterial DNA Replication and More. Annual Review of Microbiology *60*, 351-371.

Katayama, T., Akimitsu, N., Mizushima, T., Miki, T., and Sekimizu, K. (1997). Overinitiation of chromosome replication in the Escherichia coli dnaAcos mutant depends on activation of oriC function by the dam gene product. Mol Microbiol *25*, 661-670.

Kleckner, N. (1990). Regulating Tn10 and Is10 Transposition. Genetics *124*, 449-454.

Kornberg, A., and Baker, T. A. (1992). DNA replication, 2 edn (New York, W.H. Freeman).

Kouzminova, E. A., Rotman, E., Macomber, L., Zhang, J., and Kuzminov, A. (2004). RecA-dependent mutants in *Escherichia coli* reveal strategies to avoid chromosomal fragmentation. Proc Natl Acad Sci U S A *101*, 16262-16267.

Kunkel, T. A., and Erie, D. A. (2005). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.

Laurence, T. A., Kwon, Y., Johnson, A., Hollars, C. W., O'Donnell, M., Camarero, J. A., and Barsky, D. (2008). Motion of a DNA sliding clamp observed by single molecule fluorescence spectroscopy. J Biol Chem.

Liu, J., Xu, L., Sandler, S. J., and Marians, K. J. (1999). Replication fork assembly at recombination intermediates is required for bacterial growth. Proc Natl Acad Sci U S A *96*, 3552-3555.

Lobner-Olesen, A., Skovgaard, O., and Marinus, M. G. (2005). Dam methylation: coordinating cellular processes. Curr Opin Microbiol *8*, 154-160.

McCool, J. D., and Sandler, S. J. (2001). Effects of mutations involving cell division, recombination, and chromosome dimer resolution on a priA2::kan mutant. Proc Natl Acad Sci U S A *98*, 8203-8210.

McKown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988). Sequence requirements of *Escherichia coli attTn7*, a specific site of transposon Tn7 insertion. J Bacteriol *170*, 352-358.

McKown, R. L., Waddell, C. S., Arciszewska, L. A., and Craig, N. L. (1987). Identification of a transposon Tn7-dependent DNA-binding activity that recognizes the ends of Tn7. Proc Natl Acad Sci USA *84*, 7807-7811.

Miller, H. I., Kirk, M., and Echols, H. (1981). SOS induction and autoregulation of the *himA* gene for site-specific recombination in *Escherichia coli.* Proc Natl Acad Sci U S A *78*, 6754-6758.

Miller, J. H. (1992). A short course in bacterial genetics (New York, Cold Spring Harbor).

Nowosielska, A., and Marinus, M. G. (2008). DNA mismatch repair-induced double-strand breaks. DNA Repair (Amst) *7*, 48-56.

Nowosielska, A., Smith, S. A., Engelward, B. P., and Marinus, M. G. (2006). Homologous recombination prevents methylation-induced toxicity in Escherichia coli. Nucleic Acids Res *34*, 2258-2268.

Peters, J. E. (2007). Gene transfer - Gram-negative bacteria, Chapter 31. In Methods for General and Molecular Microbiology, S. L. R., ed. (Washington D. C., ASM Press), pp. 735-755.

Peters, J. E., and Craig, N. L. (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol Cell *6*, 573-582.

Peters, J. E., and Craig, N. L. (2001). Tn7 recognizes target structures associated with DNA replication using the DNA binding protein TnsE. Genes & Dev *15*, 737-747.

Peters, J. E., Thate, T. E., and Craig, N. L. (2003). Definition of the *Escherichia coli* MC4100 genome by use of a DNA array. J Bacteriol *185*, 2017-2021.

Reznikoff, W. S. (1993). The Tn5 transposon. Ann Rev Microbiol *47*, 945-963.

Robbins-Manke, J. L., Zdraveski, Z. Z., Marinus, M., and Essigmann, J. M. (2005). Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient Escherichia coli. J Bacteriol *187*, 7027-7037.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd edn (Cold Spring Harbor, NY, Cold Spring Harbor Press).

Sandler, S. J. (2000). Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. Genetics *155*, 487-497.

Sandler, S. J. (2005). Requirements for replication restart proteins during constitutive stable DNA replication in *Escherichia coli* K-12. Genetics *169*, 1799-1806.

Sandler, S. J., and Marians, K. J. (2000). Role of PriA in replication fork reactivation in Escherichia coli. J Bacteriol *182*, 9-13.

Sandler, S. J., Marians, K. J., Zavitz, K. H., Coutu, J., Parent, M. A., and Clark, A. J. (1999). dnaC mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. Mol Microbiol *34*, 91-101.

Sandler, S. J., Samra, H. S., and Clark, A. J. (1996). Differential suppression of priA2::kan phenotypes in Escherichia coli K-12 by mutations in priA, lexA, and dnaC. Genetics *143*, 5-13.

Shi, Q., Parks, A.R., Potter, B.D., Safir, I.J., Luo, Y., Forster, B.M., Peters, J.E. (2008). DNA damage differentially activates regional chromosomal loci for Tn7 transposition in *E. coli*. Genetics *In Press*.

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995). E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell *82*, 927-936.

Stellwagen, A., and Craig, N. L. (1997). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein which activates the bacterial transposon Tn7. Genetics *145*, 573-585.

Stellwagen, A., and Craig, N. L. (1998). Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. TIBS *23*, 486-490.

Su'etsugu, M., Takata, M., Kubota, T., Matsuda, Y., and Katayama, T. (2004). Molecular mechanism of DNA replication-coupled inactivation of the initiator protein in *Escherichia coli*: interaction of DnaA with the sliding clamp-loaded DNA and the sliding clamp-Hda complex. Genes Cells *9*, 509-522.

Sutera, V. A., Jr., and Lovett, S. T. (2006). The role of replication initiation control in promoting survival of replication fork damage. Mol Microbiol *60*, 229-239.

CHAPTER 5. CONCLUSIONS

The rapid growth of gene databases in recent years has presented researchers with a treasure-trove of sequences that can be analyzed by a variety easy to use of methods to address questions about a given biological system. In Chapter 2, I presented bioinformatic analysis of over 50 different Tn7-like transposons. Analysis of the Tn7-like elements lead to the discovery of a genomic island formation phenomenon resulting from the extremely specific TnsD target-site selection pathway of Tn7 transposition. I also showed that Tn7-like elements form three distinct phylogenetic groups, occupying diverse bacteria in a wide variety of ecological niches. These elements appear to contribute to the diversification and specialization of their host organisms. Examination of the genes involved in transposition of the Tn7-like elements has also generated many questions and hypotheses regarding the biology of these elements. Given the widespread nature of Tn7-like elements, it was of interest to determine how these elements spread to new hosts and functioned within such different intracellular milieus.

It was known that TnsE is the protein that is responsible for directing transposition into conjugal plasmids, but how does it recognize a minority of mobile DNA amongst a sea of other genomic DNA? A clue was offered by analysis of TnsE homolog alignments in which there appeared to be a highly conserved region within the N-terminus of these proteins that resembled the consensus β clamp binding motif that has been determined for host proteins that interact with the clamp. After further investigation, we found that TnsE does appear to interact with the β clamp and that this interaction is essential for the activity of TnsE. A more thorough examination of the requirements for

the interaction between TnsE and β lead me to propose the model for the identification of target DNAs by TnsE that was presented in Chapter 3. In this model, TnsE interacts with β while it is topologically linked to double-stranded DNA, scans along DNA until finding a gap in the duplex DNA, then releases the clamp and recruits the transposition machinery (Figure 3.17). In addition to examining the functional role of the interaction between TnsE and β, I identified a side-effect of the interaction that leads to an observable phenotype that is consistent with disruption of normal exchange of proteins that bind to the processivity factor, activation of the SOS DNA damage response pathway. The observation that mutation or over-expression of TnsE is capable of activating the SOS response provided us with a new tool to analyze interactions with the β clamp, and may be used to analyze coordination of proteins that interact with β.

I found that over-expression and mutation of *dnaN*, the gene which produces β, can increase the TnsE-mediated transposition frequency. The β clamp is used by many proteins and is absolutely essential for many processes, so changes to β often display many varying phenotypes. It will be interesting to see if it is possible to isolate mutations in *dnaN* that affect TnsE-mediated transposition, yet do not have any measurable affect on normal DNA metabolism.

Identification of an interaction between TnsE and β opens up many avenues of future research and reveals a novel mechanism that may be used by other transposons for regulation of transposition or identification of target-sites. Tn7 may also become a useful tool for understanding coordination of proteins on the β clamp. In Chapter 4, I explored some of the ways in which TnsE-mediated transposition can inform our understanding of the processes

utilized by organisms to maintain the integrity of their genomes. I analyzed mutations of host genes involved in DNA metabolism that affected TnsABC+E transposition. By monitoring TnsE-mediated transposition in selected mutant backgrounds, I observed effects that may be uniquely identified by TnsE and may lead to a better understanding the molecular mechanisms that are involved in the maintenance of the bacterial chromosome, particularly as they relate to β clamp usage in these pathways.